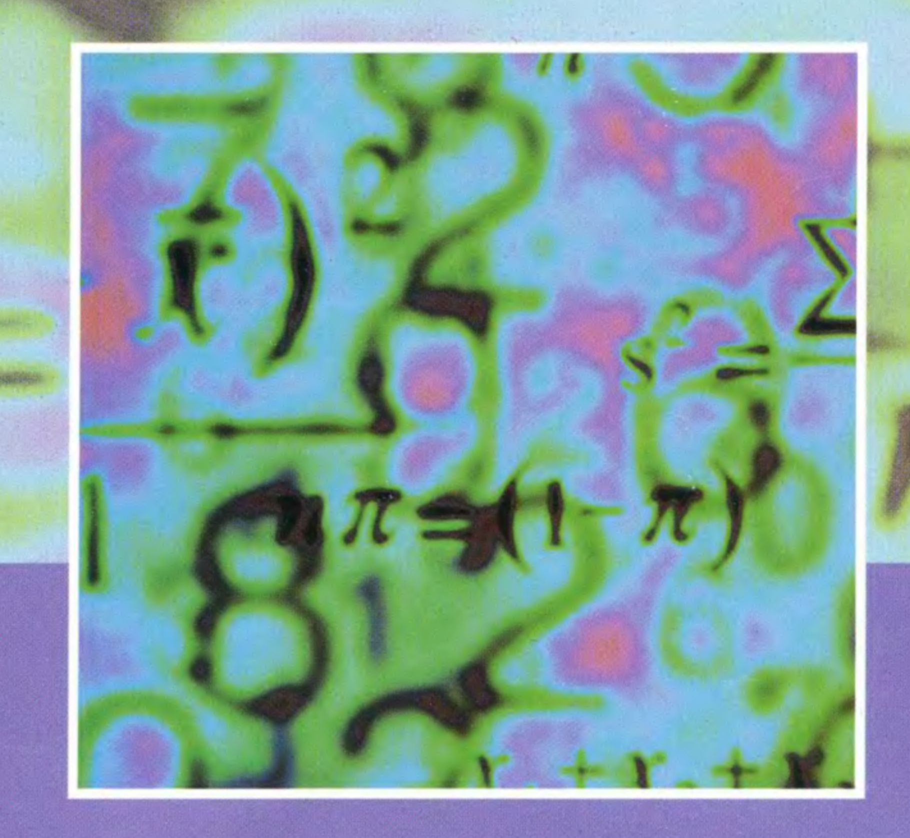


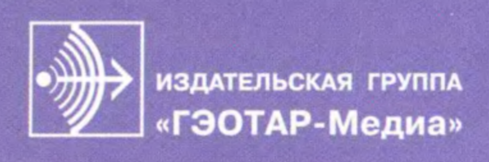
Авива Петри, Кэролайн Сэбин

Наглядная МЕДИЦИНСКАЯ СТАТИСТИКА

Перевод с английского
под редакцией В.П. Леонова

2-е издание, переработанное и дополненное





ИЗДАТЕЛЬСКАЯ ГРУППА
«ГЭОТАР-Медиа»

Medical Statistics at a Glance

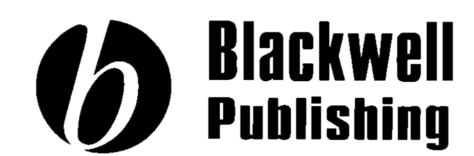
Aviva Petrie

Head of Biostatistics Unit and Senior Lecturer
Eastman Dental Institute
University College London
256 Grays Inn Road
London WC1X 8LD and
Honorary Lecturer in Medical Statistics
Medical Statistics Unit
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT

Caroline Sabin

Professor of Medical Statistics and Epidemiology
Department of Primary Care and Population Sciences
Royal Free and University College Medical School
Rowland Hill Street
London NW3 2PF

Second edition

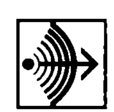


Авива Петри, Кэролайн Сэбин

Наглядная МЕДИЦИНСКАЯ СТАТИСТИКА

Перевод с английского
под редакцией В.П. Леонова

2-е издание, нереработанное и дополненное



Москва
Издательская группа «ГЭОТАР-Медиа»
2009

УДК 615.036.2=111=03.161.1(075.8)
ББК 51.1я73
П30

Рекомендовано Учебно-методическим объединением по медицинскому и фармацевтическому образованию вузов России в качестве учебного пособия для студентов медицинских вузов

Рецензенты:

Власов Василий Викторович — д-р мед. наук, проф.
Комарова Марина Валериевна — канд. биол. наук, доцент кафедры радиотехники и медицинских диагностических систем Самарского государственного аэрокосмического университета

Петри А., Сэбин К.
П30 Наглядная медицинская статистика / А. Петри, К. Сэбин ; пер. с англ. под ред. В.П. Леонова. — 2-е изд., перераб. и доп. — М. : ГЭОТАР-Медиа, 2009. — 168 с. : ил.
ISBN 978-5-9704-0914-5

Цель второго издания книги, как и первого, — донести до читателя основные понятия и принципы медицинской статистики, которые довольно широко применяют зарубежные медики и биологи. Книга содержит не только необходимую теоретическую часть, но и в доступной форме даёт практическое описание того, как можно применять статистические методы в реальных клинических исследованиях. Для освоения изложенного материала читателю вполне достаточно знаний по математике в объёме школьного курса.

Учебное пособие предназначено для студентов и аспирантов медицинских вузов, биологических факультетов университетов, врачей, исследователей-клиницистов и всех тех, кто интересуется применением статистики в медицине и биологии.

УДК 615.036.2=111=03.161.1(075.8)
ББК 51.1я73

ISBN 978-5-9704-0914-5

Содержание

Предисловие редактора к изданию на русском языке

Статистика — часть математики. Название этой отрасли знания происходит от греческого слова «матейн» (*mathein*) — учиться, познавать. Так что с полным основанием всех, кто в древнем мире изучал основы медицины, можно было называть математиками. Более того, древние греки считали науку, познание (*mathema*) и математику (*mathematike*) синонимами. Если основной метод математики — абстрагирование, то в медицине, наоборот, лечение конкретного пациента требует учёта частностей, деталей, поскольку «лечат не болезнь, а больного». Именно здесь наиболее чётко проявляется «водораздел» между медициной и математикой. Не случайно многие, кто поступает в медицинские вузы, не в ладах с математикой. И ничего плохого в этом нет. Это лишь свидетельствует о разном складе ума у людей и о востребованности разных систем мышления в обществе. В предисловии авторы говорят, что «Наглядная медицинская статистика» «...предназначена для медиков-интернов, медицинских исследователей, аспирантов биомедицинских дисциплин и персонала фармацевтической промышленности». Общеизвестно, что статистика — прекрасный инструмент, используемый при обобщении множества наблюдений. Ясно и то, что студенты II–III курса медицинского вуза накопить такие наблюдения ещё не в состоянии, поэтому-то не случайно авторы делают акцент именно на этой аудитории.

В 2006 г. при подготовке тематического номера «Международного журнала медицинской практики» посвящённого обучению медиков статистике, мы проанализировали анкету, содержавшую 26 вопросов [91]. Анализ результатов анкетирования показал, что статистический инструментарий прежде всего необходим исследователям, занятым научной работой. Современная статистическая наука и технология по своей сложности и трудности не уступает сложности и трудности разнообразных медицинских специальностей, поэтому учёному в области медицины и биологии необходимо освоить статистические технологии на уровне, достаточном для того, чтобы адекватно понимать использованный метод и в дальнейшем интерпретировать полученные результаты. Такой же точки зрения придерживаются и авторы этой книги, давая лишь основные идеи, несложные формулы и примеры применения метода. В этом основное достоинство книги, которая требует от читателя, чтобы он знал математику лишь на уровне школьного курса.

Как зарубежный, так и отечественный опыт показывает, что можно хорошо овладеть этим инструментом, если упражняться в его использовании на собственных наблюдениях. В справедливости такого утверждения автор этих строк неоднократно убеждался при подготовке и проведении выездных семинаров по биометрике. Слушатели таких семинаров обучаются на результатах предварительно выполненного нами статистического анализа их данных, что позволяет им выработать на семинаре навык адекватной интерпретации результатов такого анализа. Осенью 2007 г., занимаясь подготовкой такого семинара в Красноярске, я впервые столкнулся с тем, что организаторы семинара за полгода его подготовки не смогли собрать собственные данные для предварительного анализа. В результате профессор Ш., занимавшаяся организацией семинара, предложила обучать слушателей на чужих примерах. Очевидно, что эффективность такого семинара будет сродни результатам обучения гинеколога с использованием стоматологического кресла. Не случайно слушатели этого семинара, состоявшегося в конце января — начале февраля 2008 г., в анкетах, приведённых на сайте БИОМЕТРИКА, единодушно отметили необходимость использовать собственные данные для успешного обучения, поэтому читателю, работающему с этой книгой, следует, изучая тот или иной метод, сразу же пытаться применить его к собственным данным. Те же, у кого таких данных не будет под рукой, могут найти близкие к их специальности в библиотеке RusDASL (российская библиотека данных для изучающих биостатистику) [136], где в формате EXCEL представлены материалы из различных областей медицины и биологии.

В последние 50 лет статистику в медицине часто использовали весьма некорректно, иногда лишь как средство «онаучивания» полученных результатов исследования. Так, в 1955 г. в книге А.Я. Боярского «Статистические методы в экспериментальных медицинских исследованиях» говорилось следующее: «Уже беглое ознакомление с состоянием дела показывает, что статистическая обработка экспериментальных данных является наиболее слабым местом во многих исследованиях. ... Трудно требовать от медика, чтобы он,

наряду со знаниями в своей области, был в то же время достаточно компетентен, скажем, в радиотехнике для конструирования аппаратуры, улавливающей биотоки, или в статистике для нахождения наиболее правильных методов статистической обработки своих экспериментальных данных. И подобно тому, как медику, несомненно, приходится обращаться за содействием к радиотехнику, для правильной статистической обработки экспериментальных данных нередко приходится обращаться к специалисту-статистику. Так или иначе, но бесспорным фактом являются и недостаточная вооружённость медиков статистическими знаниями, и недостаточно высокий научный уровень статистической методики в большинстве их экспериментальных работ».

Проведённый нами анализ огромного количества статей, книг и диссертаций по медицине и биологии показал, что и спустя 50 лет мало что изменилось. По-прежнему более половины публикаций содержит многочисленные ошибки использования статистики, которые делают сомнительной и бездоказательной значительную часть последующих умозаключений авторов. С анализом наиболее типичных ошибок таких публикаций читатели могут познакомиться в литературе [88]. Причины устойчивости этого явления разнообразны [92]. Во-первых, большинство авторов — не профессиональные исследователи и в силу этого они не владеют необходимыми знаниями в области статистики. Во-вторых, нередко деформирована сама цель публикации. Она подчас воспринимается как средство априорного доказательства «научности» результатов, невзирая при этом на надёжность и доказательность выводов. Этому же способствует и то, что возможные негативные последствия опубликования сомнительных результатов отдалены от автора во времени и пространстве. Играет свою роль и то, что доступ широкой читательской аудитории как к самим публикациям, так и к исходным авторским данным ограничен. Не меньший вклад в это вносит и незаинтересованность редакций, выпускающих периодические издания, ректоров медицинских вузов, диссертационных советов, ВАК РФ и дирекций НИИ в повышении качества журнальных публикаций и диссертаций. Очевидно, что для изменения такого положения необходимы последовательные, но в то же время и радикальные меры.

Как известно, с середины 2006 г. диссертанты обязаны публиковать в Интернете авторефераты своих диссертаций. А в скором времени будут публиковать до защиты и всю диссертацию. Именно об этом заявил 31 октября 2007 г. первый вице-премьер Д.А. Медведев на встрече с членами ВАК и ректорами вузов, поддержав наши давние предложения, на что председатель ВАК декан биологического факультета МГУ академик Михаил Кирпичников ответил, что «в ближайшее время мы будем готовы говорить о публикации полностью диссертаций». Очевидно, что в связи со всем сказанным актуальность этого и других подобных изданий, знакомящих медиков с основами биостатистики, будет только возрастать. Этому же способствуют и возросшая доступность многочисленных статистических пакетов для исследователей, а также возможность дистанционно обучаться работе на них (http://www.biometrica.tomsk.ru/edu_1.htm). Востребованность настоящего издания обусловлена и тем, что усиливается интерес к концепции доказательной медицины, в которой статистика — один из основных инструментов.

Как и следовало ожидать, второе издание книги содержит некоторые дополнения, о которых авторы сами говорят в предисловии. Можно предполагать, что новые главы появились как результат освоения авторами новых методов, которые они посчитали актуальными и полезными. Так как в обоих изданиях отсутствуют списки рекомендуемой литературы по биостатистике, мы дополнили второе издание весьма обширной библиографией. Более полный перечень книг приведён по адресам: http://www.biometrica.tomsk.ru/razdel_2_1.htm — http://www.biometrica.tomsk.ru/razdel_2_12.htm.

Поскольку отечественная терминология в этой области ещё только складывается, возможно, некоторые термины могут показаться читателям непривычными. В перевод этого издания мы добавили некоторые примечания, конкретизирующие ряд деталей. Несомненное достоинство издания «Study of Case» — обучение на примерах, наличие обширного глоссария и детального предметного указателя. Всё это позволяет утверждать, что книга найдёт своего читателя и позволит повысить уровень знаний о статистике многим специалистам в области медицины.

Канд. тех. наук, доцент факультета информатики
Томского государственного университета,
редактор сайта БИОМЕТРИКА,
http://www.biometrica.tomsk.ru

В.П. Леонов

Предисловие

«Наглядная медицинская статистика» (*Medical Statistics at a Glance*) предназначена для медиков-интернов, медицинских исследователей, аспирантов биомедицинских дисциплин и персонала фармацевтической промышленности. Все эти лица в профессиональной практике неизбежно столкнутся с количественными результатами (своими собственными или чужими), которые потребуют критической оценки и интерпретации, а некоторые из них, конечно, будут вынуждены пройти экзамен по этой наводящей ужас статистике! Надлежащее понимание статистических концепций и методологии неоценимо для этих целей. Будучи прагматиками, мы хотели бы зажечь читателя энтузиазмом в области статистики. Наша цель — дать студенту и исследователю, как и клиницистам, которые сталкиваются со статистическими концепциями в медицинской литературе, книгу, которая логична, легко читается, исчерпывающа и имеет практическое применение.

Мы полагаем, что «Наглядная медицинская статистика» будет особенно полезна как дополнительный материал к лекциям по статистике и как руководство с соответствующими ссылками. Вместе с другими книгами серии «At a Glance» мы ведём читателя через ряд самостоятельных двух- и трёхстраничных тем, каждая из которых охватывает отдельный аспект медицинской статистики. Из опыта обучения мы выяснили и приняли во внимание трудности, с которыми столкнулись наши студенты при изучении медицинской статистики. По этой причине мы предпочли ограничить теоретическое содержание книги уровнем, достаточным для понимания включённых в неё процедур, но всётаки не затмевающим практические аспекты их выполнения.

Медицинская статистика — предмет, имеющий широкий диапазон и включающий множество разделов. Мы даём элементарное введение в основополагающие концепции медицинской статистики и руководство по наиболее часто применяемым статистическим процедурам. Эпидемиология тесно связана с медицинской статистикой, поэтому обсуждаются некоторые основные вопросы, относящиеся к разработке исследования и его интерпретации. Включены также темы, которые могут быть полезны читателю только изредка, но которые, тем не менее, необходимы во многих областях медицинских исследований: например, доказательная медицина, систематические обзоры и метаанализ, анализ временных рядов, анализ выживаемости и байесовские методы. Мы объяснили принципы, лежащие в основе этих тем, так, чтобы читатель смог понять и интерпретировать результаты, когда они встречаются в литературе.

Порядок первых 30 глав этого издания не отличается от первого издания. Большинство из них остались неизменёнными и в новом издании, некоторые же содержат незначительные отличия, представляющие собой последние результаты и взаимные ссылки, или преобразованы с учётом нового материала. Главные поправки касаются сравнительно сложных форм регрессионного анализа, которые сейчас используются гораздо шире, чем тогда, когда был написан первый вариант книги, частично потому, что соответствующее программное обеспечение стало доступнее и эффективнее, чем прежде. Мы изменили главу, посвящённую бинарным откликам (исходам) и логистической регрессии, включили новую главу об интенсивности и пуассоновской регрессии и значительно расширили первоначальную тему по статистическому моделированию, так что теперь она включает три главы: «Общие линейные модели», «Объясняющие переменные в статистических моделях» и «Актуальные вопросы в статистическом моделировании». Мы также изменили главу 41, которая описывает различные подходы к анализу сгруппированных данных, и добавили главу 42, выделив различные методы регрессии, которые можно использовать для анализа этого типа данных. Первое издание содержало краткое описание анализа временных рядов, которое мы решили не включать во второе издание, поскольку чувствовали. что оно, вероятно, было весьма ограниченным, чтобы его практически использовать, а его расширение отодвинуло бы сроки издания. Из-за этого пропуска и некоторых дополнений нумерация глав, начиная с главы 31, в первом и во втором издании различается. Большинство из них по сравнению с первым изданием либо слегка изменены, либо остались неизменёнными.

Описание каждой статистической технологии сопровождается примером, иллюстрирующим, как её использовать. Данные для примеров мы взяли из совместных исследований, в которых участвовали мы или наши коллеги, а в отдельных случаях — из ранее опубликованных статей. Где возможно, мы использовали один и тот же набор данных, чтобы приблизить к реальности технологию анализа, которая редко ограничивается единственным методом или подходом. Хотя предполагается, что формулы должны подтверждать логику подхода, помогая объяснению и пониманию, мы старались не приводить детально сложные вычисления: большинство читателей имеют доступ к компьютеру и вряд ли станут выполнять любые, даже самые простые вычисления вручную.

Мы полагаем, что для читателя особенно важно умение интерпретировать результаты вычислений на компьютере, поэтому предпочли там, где это возможно, иллюстрировать результаты фрагментами компьютерных вычислений. В отдельных случаях, когда с такой интерпретацией могут возникнуть трудности, мы включали (приложение В) и аннотированные компьютерные распечатки (листинги), содержащие анализ набора данных. Есть много статистических пакетов общего применения; для того чтобы показать читателю, как эти распечатки могут видоизменяться, мы не ограничились подробным разбором какого-либо одного конкретного пакета, а использовали три хорошо известных статистических пакета: SAS, SPSS и STATA.

В тексте книги есть обширные перекрёстные ссылки, чтобы помочь читателю связать различные процедуры. Основной набор статистических таблиц содержится в приложении А (Neave H.R. Elementary Statistical Tables Routledge, 1981, и Diem K. Documenta Geigy Scientific Tables. — 7th edn. — Oxford: Blackwell Publishing, 1970), которые среди прочих дают более полные версии, если читателю потребуется точнее рассчитать результаты вручную. Словарь терминов (приложение Г) содержит легкодоступные объяснения часто применяемой терминологии.

Известно, что одна из трудностей, с которыми сталкиваются нестатистики, — выбор подходящего метода, поэтому мы создали две схемы, которые можно использовать и для того, чтобы выбрать метод в данной ситуации, так и для лёгкого поиска этого метода. Они воспроизведены на внутренней стороне обложки.

Читателю могут быть полезными диалоговые упражнения на веб-сайте (http://www.medstatsaag.com/). Он содержит полный набор ссылок (некоторые из них непосредственно связаны с базой данных Medline), что позволяет расширить круг ссылок, указанных в книге, и обеспечить себя дополнительной полезной информацией для примеров. Читателям, желающим более подробно изучить специфические области медицинской статистики, мы рекомендуем следующие книги.

- Altman D.G. Practical Statistics for Medical Research. — London: Chapman and Hall, 1991.
- Armitage P., Berry G., Matthews J.F.N. Statistical Methods in Medical Research. — 4th edn. — Oxford: Blackwell Science, 2001.
- Pocock S.J. Clinical Trials: A Practical Approach. — Chichester: Wiley, 1983.

Мы чрезвычайно благодарны Марку Гилторпу и Джонатану Стерну за ценные комментарии и предложения, касающиеся ряда аспектов во втором издании, а также Ричарду Моррису, Фионе Ламп, Шаку Хаджату и Абулу Басару за обсуждение первого издания. Мы благодарим всех, кто помогал нам искать данные для примеров. Естественно, мы несём полную ответственность за любые ошибки в тексте или примерах. Мы также благодарим Майка, Джеральда, Нину, Эндрю и Карен, которые терпеливо, хладнокровно переносили нашу увлечённость, работой над первым изданием и пережили вместе с нами все испытания и несчастья, сопровождавшие издание второго.

Авива Петри,
Каролина Сэбин,
Лондон

Список сокращений

ВКК — внутриклассовый коэффициент корреляции
ДИ — доверительный интервал
ДМ — доказательная медицина
ИБС — ишемическая болезнь сердца
ИМ — инфаркт миокарда
МНК — метод наименьших квадратов
ОЛМ — обобщённая линейная модель
ОМП — оценка максимального правдоподобия
ОНК — оценка наименьших квадратов
ООУ — обобщённые оценки уравнения
ОР— относительный риск
ППК — площадь под кривой
ПРК — первичное послеродовое кровотечение
РКИ — рандомизированные контролируемые испытания
СМП — статистика отношения правдоподобия
СПИД — синдром приобретённого иммунодефицита человека
УДВВ — уровень давления в воротной вене
ЦМВ — цитомегаловирусная инфекция
ЧБНЛ — число больных, которое вам необходимо лечить
ШФА — шкала физической активности
CABG — операция шунтирования коронарной артерии
CMV — цитомегаловирус
df — «степени свободы»
HAART — высокоактивное антиретровирусное лечение
HHV-8 — вирус герпеса человека
HRT — терапия замены гормона в постменопаузе
FEV_1 — объём форсированного выдоха в 1 с
IVF — экстракорпоральное оплодотворение
LR — отношение правдоподобия
ROC — кривая операционной характеристики диагностического теста
SD — стандартное отклонение
SEM — стандартная ошибка среднего

1 Типы данных

ДАННЫЕ И СТАТИСТИКА[1]

Цель большинства исследований состоит в сборе данных, которые впоследствии помогают получить информацию относительно какой-либо области исследования. Данные всегда основаны на наблюдениях одной или нескольких переменных; термин переменная означает количественный показатель, способный изменяться. Например, мы можем собрать основную клиническую и демографическую информацию о пациентах со специфической болезнью. Интересующими нас переменными могут быть пол, возраст и рост больного.

Обычно мы получаем данные из выборки индивидуумов, представляющих популяцию — группу индивидуумов, которая представляет для нас интерес. Цель состоит в том, чтобы сгруппировать эти данные и извлечь полезную информацию. Статистика использует различные методы, например, сбор данных, их обобщение, анализ и подведение итогов, основанных на полученных сведениях.

Существуют различные формы данных. Прежде чем решить, какой статистический метод окажется наиболее подходящим для конкретного случая, мы должны знать, к какому типу данных относится каждая переменная. Все переменные, результирующие показатели, можно разделить на два типа: категориальный (качественный) или числовой (количественный) (рис. 1-1).

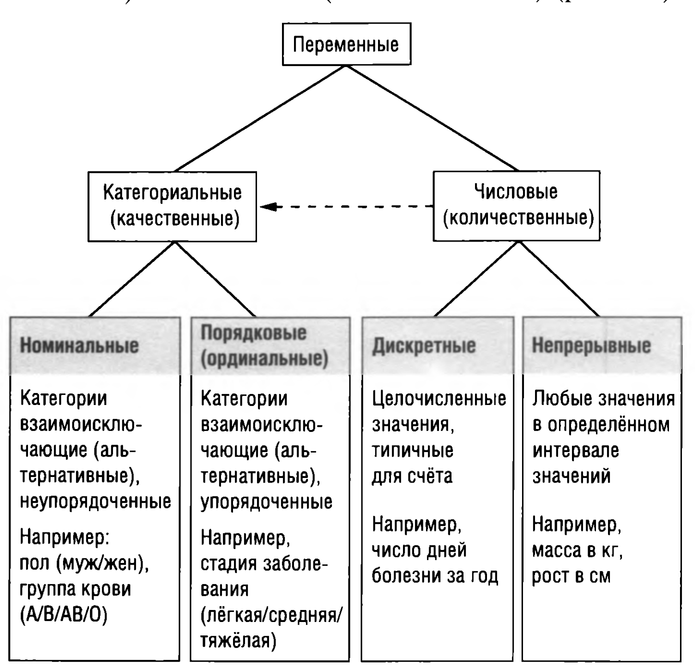


Рис. 1-1. Различные типы переменных (схема).

КАТЕГОРИАЛЬНЫЕ (КАЧЕСТВЕННЫЕ) ДАННЫЕ

Встречаются тогда, когда индивидуум может принадлежать только к одной из множества категорий переменных.

- Номинальные данные. Категории не упорядочиваются, а просто имеют названия. Например, группа крови (А, В, АВ и О) и семейное положение (замужем, вдова, не замужем и т.д.). В этом случае нет оснований полагать, что быть замужем лучше (или хуже), чем быть не замужем!
- Ординальные (ранговые, порядковые) данные. В некоторых случаях категории (градации, уровни) упорядочиваются. Например, стадии болезни (заболевание в запущенной стадии, средняя, лёгкая форма болезни или её отсутствие) и степень боли (сильная, умеренная, слабая, отсутствие боли).

Категориальная (качественная) переменная — бинарная или дихотомическая переменная, когда имеются только две возможные категории. Например, «да/нет», «умер/жив» или «пациент болен/пациент здоров».

ЧИСЛОВЫЕ (КОЛИЧЕСТВЕННЫЕ) ДАННЫЕ

Встречаются, когда переменная имеет некоторую числовую величину (значение). Мы можем подразделить числовые данные на два типа.

- Дискретные данные. Переменная может принимать только определённые числовые значения. Часто ведётся подсчёт количества событий, например, количество посещений врача в год или количество заболеваний человека за последние пять лет.
- Непрерывные данные. Нет никаких ограничений в отношении данных, которые переменная может принимать, например, масса тела или рост.

РАЗЛИЧИЕ МЕЖДУ ТИПАМИ ДАННЫХ

В зависимости от того, оказываются ли данные категориальными или числовыми, используют различные статистические методы. Казалось бы, различия между категориальными и числовыми данными и так понятны, но в некоторых случаях отличие становится не совсем ясным. Например, когда у нас есть переменная с множеством установленных категорий (боль может иметь семь категорий), могут возникнуть трудности, как отличить её от дискретной числовой переменной. Различие между дискретными и непрерывными числовыми данными может быть даже менее понятным, хотя в общем на результатах большинства исследований это не отразится. Возраст — пример переменной, которую часто трактуют как дискретную, хотя её нужно отнести к непрерывным. Обычно мы ссылаемся на «возраст в последний день рождения», нежели на «возраст по состоянию на сегодняшний день», и поэтому женщине, которая сообщает, что ей 30, может быть 30 лет и несколько дней или почти 31 год.

Не торопитесь вначале записывать числовые данные как категориальные, поскольку часто теряется важная информация. Гораздо проще преобразовать числовые данные в категориальные данные сразу же, как только они будут собраны.

[1] Детальное изложение особенностей различных типов данных [127]. Прим. переводчика.

ПРОИЗВОДНЫЕ (ВТОРИЧНЫЕ) ДАННЫЕ

Существует множество других типов данных в области медицины. Они могут включать в себя следующее.

- Проценты. Могут возникать при рассмотрении вопроса относительно улучшения состояния больного во время лечения, например состояние больного (объём форсированного выдоха в 1 с, FEV_1) может улучшиться на 24% после лечения новым препаратом. В этом случае имеет значение степень улучшения, а не абсолютные данные о состоянии пациента.
- Пропорции или отношения. Иногда встречается два варианта пропорций или отношений. Например, индекс массы тела (индекс Кетле): отношение массы тела индивидуума (кг) к квадрату роста, таким образом проводится оценка, превышает ли масса норму или, наоборот, недостаточна.
- Интенсивность. Относительная частота заболеваний, где количество заболеваний делят на общее число лет, в течение которых вели наблюдения за пациентами в этом исследовании (см. главу 31), общепринята при эпидемиологическом исследовании (см. главу 12).
- Метки, оценки. Произвольные значения, или метки, используют в том случае, когда невозможно измерить количество. Например, ряд вопросов на ответы относительно качества жизни можно суммировать для того, чтобы дать полную оценку относительно качества жизни у каждого индивидуума[1].

Все эти переменные можно рассматривать в большинстве исследований как непрерывные переменные. В данном случае переменную можно будет сконструировать, если использовать более чем одну величину (например, числитель и знаменатель для процента), важно при этом регистрировать все используемые значения. Например, состояние больного улучшилось на 10% после лечения, данное улучшение может иметь различную клиническую значимость в зависимости от того, в каком состоянии находился больной до лечения.

ЦЕНЗУРИРОВАННЫЕ ДАННЫЕ

Мы можем рассмотреть цензурированные данные на следующих примерах.

- Если мы проводим лабораторные измерения, используя прибор, который может обнаружить значения только выше некоторого предельного уровня, тогда любая величина ниже этого уровня не будет обнаружена. Например, вирус, уровень обнаружения которого ниже предела, часто рассматривается как «необнаруженный», при том что на самом деле он может находиться в образце.
- Мы можем столкнуться с цензурированными данными, например, когда некоторые больные из группы исследуемых отстраняются от испытания до окончания исследований. Этот тип данных подробно обсуждается в главе 44.

[1] Такие числовые метки могут быть получены при анализе наборов качественных и количественных признаков. Многие из них — результаты применения ряда многомерных методов, таких, как дискриминантный анализ, корреспондентский анализ и т.д., не рассмотренных в этой книге. Автор [49] даже ввёл новый термин «патометрия» для обозначения технологии получения количественных меток, характеризующих степень патологии. Прим. переводчика.

2 Ввод данных

При проведении какого-либо исследования вам почти всегда необходимо будет вводить данные в компьютерный пакет прикладных программ. Компьютер — неоценимая вещь, с его помощью можно проверить правильность данных, ускорить сбор данных и анализа, гораздо проще проверять ошибки, проводить графические подсчёты данных и новые переменные. Стоит потратить некоторое время на планирование ввода данных — и на последней стадии это сэкономит ваше время и усилия.

ФОРМАТЫ ДЛЯ ВВОДА ДАННЫХ

Существует несколько способов ввода данных и сохранения их в компьютере. Большинство статистических пакетов позволяют сразу же вводить данные. Однако существуют ограничения, а именно: вы не сможете переносить данные из одного пакета в другой. Простейшая альтернатива — сохранять данные либо в электронной таблице, либо в пакете баз данных. К сожалению, их статистические процедуры часто ограничены, и обычно возникает необходимость вводить данные в статистический пакет, чтобы провести исследования.

Наиболее гибкий подход состоит в том, чтобы сохранять ваши данные как ASCII (American Standard Code for Information Interchange — стандартный код информационного обмена США) или в текстовом файле. Данные в ASCII-формате могут читаться большинством пакетов. Формат ASCII состоит просто из текста, который вы можете читать с компьютера. Обычно каждая переменная в файле отделяется от следующей каким-нибудь разделителем, часто пробелом или запятой. Такой формат известен как свободный формат.

Самый простой способ ввода данных в ASCII-формате — печатать данные прямо в этом формате, используя текстовый редактор либо иной редакторский пакет. В качестве альтернативы данные, находящиеся в пакете электронных таблиц (Excel), могут быть сохранены в текстовом формате[1].

Используя любой подход, при исследовании принято, чтобы каждая строка данных соответствовала отдельному индивидууму и каждая колонка соответствовала переменной, хотя может возникнуть необходимость в продолжении последовательных рядов, в случае если на каждого индивидуума собрано большое количество переменных.

ПЛАНИРОВАНИЕ ВВОДА ДАННЫХ

При сборе данных вам необходимо будет использовать форму или анкету для их фиксации. Если они хорошо разработаны, то могут сократить работу, которую необходимо выполнить при вводе данных. В общем, эти формы/анкеты включают ряд «ячеек», в которые заносятся данные — обычно имеется отдельная «ячейка» для каждого возможного числового ответа.

[1] Одна из наиболее часто встречающихся ошибок начинающих исследователей заключается в том, что они вводят данные в таблицы EXCEL в текстовом формате и затем при импорте этих таблиц в какой-либо статистический пакет, например пакет STATISTICA или SPSS, испытывают сложности при попытке анализа этих данных как числовых. Прим. переводчика.

КАТЕГОРИАЛЬНЫЕ ДАННЫЕ

С нечисловыми данными могут возникнуть проблемы при занесении их в некоторые статистические пакеты, поэтому вам необходимо назначить числовые коды категориальным данным, прежде чем вводить данные в компьютер. Например, вы можете выбрать следующие коды: 1, 2, 3 и 4 категориям «нет боли», «лёгкая боль», «средняя боль» и «сильная боль» соответственно. Эти коды могут быть добавлены к формам при сборе данных. Для бинарных данных, например, ответов «да/нет», очень удобно установить код 1 (например, для «да») и 0 (для «нет»).

- Переменные с единственным альтернативным вариантом ответа. Существует только один возможный ответ на вопрос. Например, на вопрос «Умер ли пациент?» невозможно ответить и «да», и «нет».
- Переменные с несколькими альтернативами ответа. Возможен более чем один ответ. Например, на вопрос: «Каковы симптомы болезни у пациента?» — можно перечислить несколько симптомов. Существует два способа обработки этих данных в зависимости от того, какую из двух следующих ситуаций применить.
  - ✧ Существует несколько возможных симптомов и многие из них присутствуют у пациента. Можно создать ряд различных бинарных переменных, все зависит от того, ответит ли больной «да» или «нет» относительно возможных симптомов. Например, был ли у него кашель, болело ли горло.
  - ✧ Существует огромное количество возможных симптомов, но у пациента могут быть только некоторые из них. Можно создать ряд различных номинальных переменных, каждая из которых позволит определить наличие того или иного симптома у больного. Например, какой симптом возник первым, какой — вторым и т.д. Вы заранее должны определить максимальное количество симптомов, которые, вы полагаете, могут быть у больного.

ЧИСЛОВЫЕ ДАННЫЕ

Числовые данные должны быть введены с той же самой точностью, с которой были проведены измерения, и единица измерения должна быть едина для всех наблюдений данной переменной. Например, масса должна быть записана в килограммах или в граммах, но не попеременно то в килограммах, то в граммах.

МНОЖЕСТВЕННЫЕ ФОРМЫ НА ОДНОГО БОЛЬНОГО

Иногда информацию собирают на одного и того же больного более чем в одном случае (наблюдении). Важно отметить, что должен существовать уникальный идентификатор (например, порядковый номер), принадлежащий только одному человеку в данном наблюдении, который предоставит вам возможность соединить все данные, собранные на одного человека при исследовании.

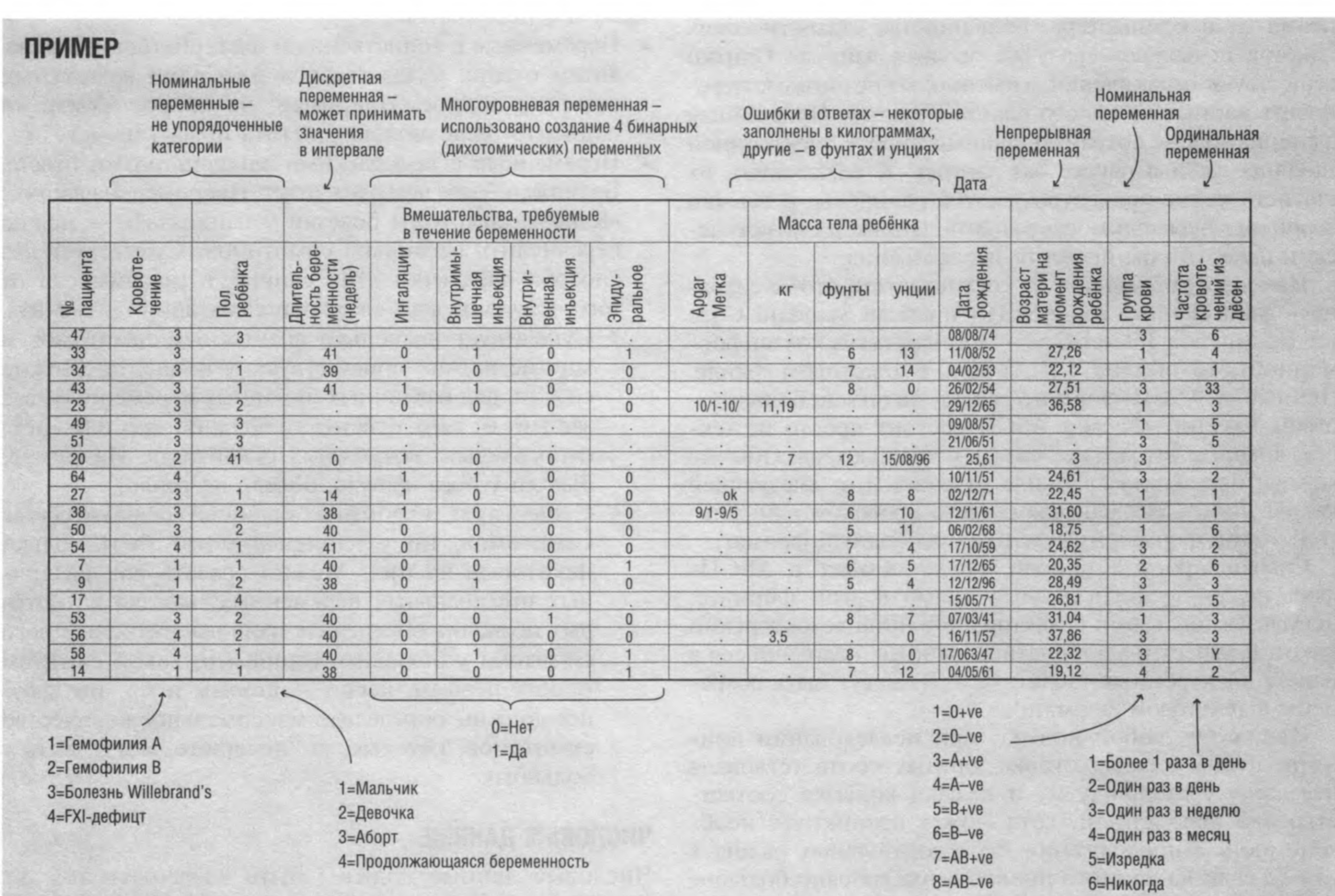
ПРОБЛЕМЫ С ДАТАМИ И ПЕРИОДАМИ

Даты и периоды необходимо вводить последовательно, например: либо день/месяц/год, либо месяц/день/год, но всегда в одном и том же порядке. Важно установить, какой формат может читаться в данном статистическом пакете.

КОДИРОВАНИЕ ОТСУТСТВУЮЩИХ (ПРОПУЩЕННЫХ) ДАННЫХ

Вам следует определить, что вы будете делать с отсутствующими данными, прежде чем вводить информацию. В большинстве случаев вы будете вынуждены использовать какой-нибудь символ для недостающих данных. Статистические пакеты предлагают для этого различные способы. Некоторые пакеты используют специальные символы (например, точку или звездочку) для обозначения пропущенных данных, принимая это во внимание во время анализа, другие требуют от вас ввести свой собственный код для обозначения отсутствующих данных (обычно используют знаки 9, 999 или −9999). Выбранное значение должно быть одно для всех переменных, и его невозможно использовать для другой переменной. Например, при вводе категориальной переменной с четырьмя категориями (коды 1, 2, 3 и 4) вы можете выбрать цифру 9 для недостающих данных. Однако для переменной «возраст ребёнка» будет необходимо выбрать другой код, например, «−9». Более подробно отсутствующие данные рассмотрены в главе 3.

ПРИМЕР



№ пациента	Кровотечение	Пол ребёнка	Длительность беременности (недель)	Вмешательства, требуемые в течение беременности: Ингаляции	Внутримышечная инъекция	Внутривенная инъекция	Эпидуральное	Apgar Метка	Масса тела ребёнка: кг	фунты	унции	Дата рождения	Возраст матери на момент рождения ребёнка	Группа крови	Частота кровотечений из дёсен
47	3	3	.	.	.	.	.	.	.	.	.	08/08/74	.	3	6
33	3	.	41	0	1	0	1	.	.	6	13	11/08/52	27,26	1	4
34	3	1	39	1	0	0	0	.	.	7	14	04/02/53	22,12	1	1
43	3	1	41	1	1	0	0	.	.	8	0	26/02/54	27,51	3	33
23	3	2	.	0	0	0	0	10/1-10/	11,19	.	.	29/12/65	36,58	1	3
49	3	3	.	.	.	.	.	.	.	.	.	09/08/57	.	1	5
51	3	3	.	.	.	.	.	.	.	.	.	21/06/51	.	3	5
20	2	41	0	1	0	0	.	.	7	12	15/08/96	25,61	3	3	.
64	4	.	.	1	1	0	0	.	.	.	.	10/11/51	24,61	3	2
27	3	1	14	1	0	0	0	ok	.	8	8	02/12/71	22,45	1	1
38	3	2	38	1	0	0	0	9/1-9/5	.	6	10	12/11/61	31,60	1	1
50	3	2	40	0	0	0	0	.	.	5	11	06/02/68	18,75	1	6
54	4	1	41	0	1	0	0	.	.	7	4	17/10/59	24,62	3	2
7	1	1	40	0	0	0	1	.	.	6	5	17/12/65	20,35	2	6
9	1	2	38	0	1	0	0	.	.	5	4	12/12/96	28,49	3	3
17	1	4	.	.	.	.	.	.	.	.	.	15/05/71	26,81	1	5
53	3	2	40	0	0	1	0	.	.	8	7	07/03/41	31,04	1	3
56	4	2	40	0	0	0	0	.	3,5	.	0	16/11/57	37,86	3	3
58	4	1	40	0	1	0	1	.	.	8	0	17/063/47	22,32	3	Y
14	1	1	38	0	0	0	1	.	.	7	12	04/05/61	19,12	4	2

Рис. 2-1. Часть электронной таблицы, в которой показаны собранные данные на примерах 64 женщин с наследственными беспорядочными кровотечениями.

Эта часть исследования показывает, как влияют наследственные беспорядочные кровотечения на беременность и роды. Данные были собраны при исследовании 64 женщин, зарегистрированных в одном и том же Центре гемофилии в Лондоне. Женщин опрашивали относительно их кровотечений и их первой беременности (или их текущей беременности, если они были беременны в первый раз на день интервьюирования). На рис. 2-1 представлены данные, которые были собраны при исследовании небольшого количества женщин после того, как данные были введены в электронную таблицу, но до того, как их проверили на наличие ошибок. Внизу на рис. 2-1 приведена кодовая схема для категориальных переменных. Каждый ряд отведён для отдельного пациента; каждая колонка для отдельной переменной. В случае если женщина всё ещё беременна, возраст женщины высчитывался на день рождения младенца. Данные, касающиеся живорождённых детей, описаны в главе 37.

Данные любезно предоставлены Dr R.A. Kadir, University Department of Obstetrics and Gynaecology, and Professor C.A. Lee, Haemophilia Centre and Haemostasis Unit, Royal Free Hospital, London.

3 Проверка ошибок и выбросов

При любом исследовании всегда есть опасность допустить ошибки при наборе данных либо вначале, при измерениях, либо при сборе, переписывании и вводе данных в компьютер. Довольно трудно избежать этих ошибок. Однако можно сократить количество опечаток и описок путём тщательной проверки данных, как только они будут введены. Даже бегло просмотрев таблицу, можно обнаружить очевидные ошибки. В этой главе мы предлагаем ряд подходов, которые можно использовать при проверке данных.

ОПЕЧАТКИ

Опечатки — самые распространённые ошибки при вводе данных. Если количество данных невелико, вы можете сравнить уже напечатанные с оригинальными, просто просмотрев их, и проверить, нет ли ошибок. Однако при большом объёме данных на это потребуется слишком много времени. Можно ввести данные дважды и сравнить их при помощи компьютерной программы. Любые различия между двумя вариантами будут обнаружены. Хотя не исключено, что одна и та же ошибка может быть допущена в обоих случаях или данные в форме/анкете неправильные, но по крайней мере это сводит к минимуму количество ошибок. Недостаток этого метода заключается в том, что приходится дважды вводить данные, а это может повлечь большие затраты денег и времени.

ПРОВЕРКА ОШИБОК[1]

- Категориальные данные. Относительно легко проверить категориальные данные, так как отклики на каждую переменную (переменная отклика) могут принимать только одно из ограниченного ряда значений, поэтому данные, которые недопустимы, должны считаться ошибочными.
- Числовые (количественные) данные. При вводе числовых данных достаточно просто поменять местами цифры или не туда поставить десятичную запятую — и данные искажены, поэтому их довольно трудно проверить, но и здесь надо попытаться устранить ошибки. Числовые данные можно проверить по размаху, т.е. задать верхние и нижние ограничения для каждой переменной. Если величина находится за пределами этого интервала, то она не используется при дальнейшем исследовании.
- Даты. Часто трудно проверить точность дат, хотя иногда вам следует знать, что в определённый период времени данные могут выпадать (исчезать). Даты необходимо проверять хотя бы ради того, чтобы удостовериться, что они действительны. Например, 30 февраля не существует, как и не может быть в месяце больше 31 дня, а в году — больше 12 месяцев. Можно применять и некоторые логические проверки. Например, дата рождения больного должна соответствовать его возрасту, больной должен родиться до начала исследования (по крайней мере, в большинстве исследований). Кроме того, больной, который умер, не может осуществлять последующие визиты!

Во всех проверках величина должна быть исправлена только в том случае, если ошибка очевидна. Не следует менять данные только потому, что они выглядят необычными.

ОБРАБОТКА ПРОПУЩЕННЫХ ДАННЫХ

Всегда может случиться так, что некоторых данных не будет. Если не хватает большого объёма информации, результаты анализа вряд ли будут надёжны. Необходимо выяснить, почему сведения отсутствуют. Если нет данных для какой-то одной переменной и/или в отдельной подгруппе индивидуумов, это может указывать на то, что переменная не используется или никогда не была измерена для этой подгруппы. Тогда эту переменную надо исключить из исследования по данной группе индивидуумов. Может оказаться, что эти данные просто ещё в работе, на бумаге и пока не введены!

ВЫБРОСЫ (АНОМАЛЬНЫЕ ЗНАЧЕНИЯ)

Что такое выбросы?

Выбросы — наблюдения, которые отличаются от главной группы данных и несовместимы с остальными. Эти данные могут быть подлинными наблюдениями с очень экстремальными величинами переменной. Однако они могут появиться также в результате опечаток и в этом случае любые данные, вызывающие подозрение, должны быть проверены. Важно выяснить, имеются ли выбросы в наборе данных, так как они могут в значительной степени повлиять на результаты некоторых исследований.

К примеру, женщина, у которой рост 2,1 м, вероятнее всего воспринималась бы как выброс в большинстве наборов данных. Однако, хотя очевидно, что эта величина довольно высокая по сравнению с обычным ростом женщин, эти данные могут быть подлинными, так как эта женщина может быть просто очень высокой. В этом случае надо исследовать это наблюдение и дальше, может быть, проверить другие её показатели, такие, как возраст и масса, прежде чем принимать решение относительно истинности этой величины.

[1] Достаточно эффективно при небольших и средних объёмах введённых данных использовать последовательные сортировки наблюдений (желательно ещё при вводе в таблицы EXCEL), по отдельным признакам (столбцам). В этом случае ошибочные значения будут размещаться в первой или в последней (в зависимости от типа сортировки — по возрастанию или уменьшению значений) строке таблицы. Кроме того, отчётливо будет видно различие между ошибочным значением и ближайшим к нему. Прим. переводчика.

И только в том случае, если стало очевидно, что эти данные неверны, следует изменить значение.

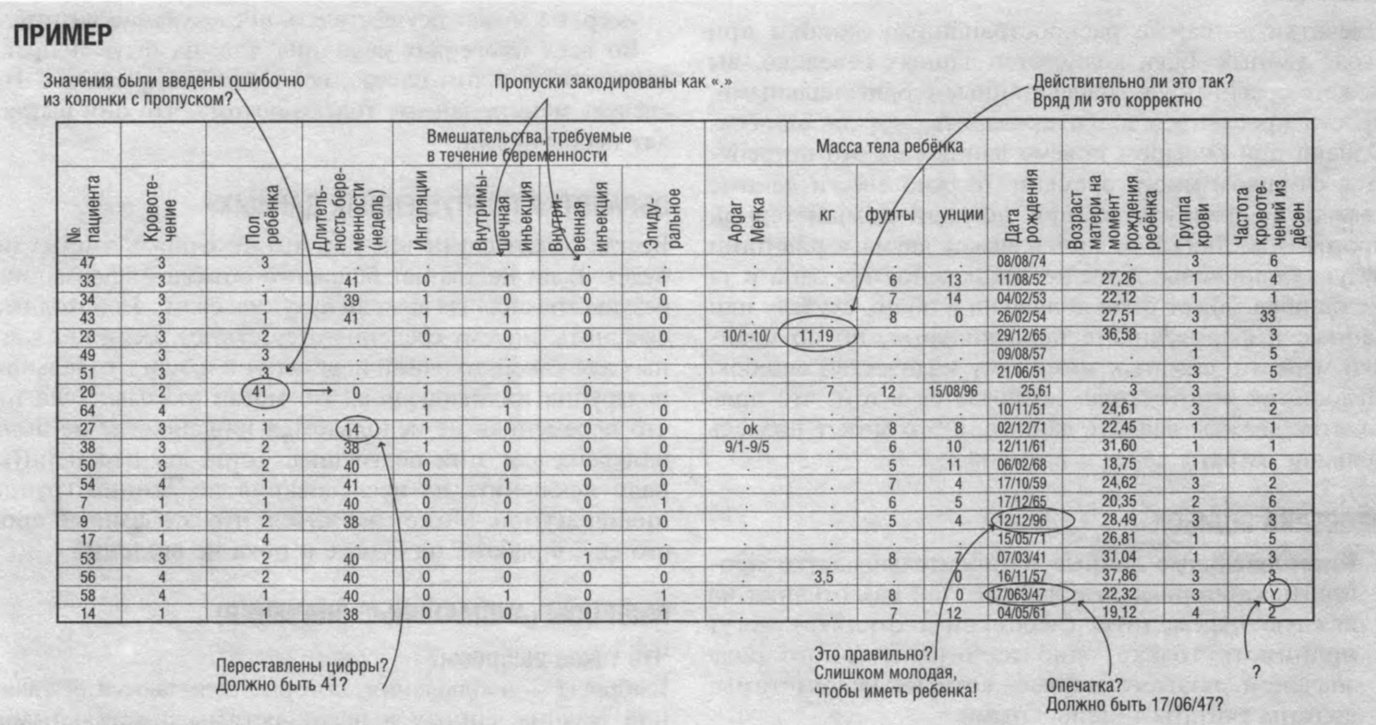
Проверка выбросов

Самый простой метод состоит в том, чтобы во время набора данных проверять их глазами. Это приемлемо, если количество наблюдений не слишком большое и потенциальный выброс намного ниже или выше, чем остальная часть данных. Проверка по интервалу изменения также должна идентифицировать возможные выбросы. В качестве альтернативы данные могут быть представлены иным способом — выбросы будут идентифицированы на гистограммах и диаграммах рассеяния (см. также главу 29 с дискуссией о выбросах в регрессионном анализе)[1].

Обращение с выбросами

Нельзя убрать индивидуума из анализа только потому, что его/её данные выше или ниже, чем должны быть. Однако включение выбросов может повлиять на результаты, когда используются какие-нибудь статистические методы. Самый простой метод состоит в том, чтобы повторить анализ как с включёнными, так и с исключёнными данными. Если результаты окажутся одинаковыми, то в этом случае выбросы не окажут большого влияния на результаты. Однако, если результаты сильно отличаются, следует применить соответствующие методы, при которых выбросы не повлияют на исследование данных. Они включают в себя применение преобразований и непараметрических критериев.

ПРИМЕР



				Вмешательства, требуемые в течение беременности					Масса тела ребёнка						
№ пациента	Кровотечение	Пол ребёнка	Длительность беременности (недель)	Ингаляции	Внутримышечная инъекция	Внутривенная инъекция	Эпидуральное	Apgar Метка	кг	фунты	унции	Дата рождения	Возраст матери на момент рождения ребёнка	Группа крови	Частота кровотечений из дёсен
47	3	3	.	.	.	.	.	.	.	.	.	08/08/74	.	3	6
33	3	.	41	0	1	0	1	.	.	6	13	11/08/52	27,26	1	4
34	3	1	39	1	0	0	0	.	.	7	14	04/02/53	22,12	1	1
43	3	1	41	1	1	0	0	.	.	8	0	26/02/54	27,51	3	33
23	3	2	.	0	0	0	0	10/1-10/	11,19	.	.	29/12/65	36,58	1	3
49	3	3	.	.	.	.	.	.	.	.	.	09/08/57	.	1	5
51	3	3	.	.	.	.	.	.	.	.	.	21/06/51	.	3	5
20	2	41	0	1	0	0	.	.	7	12	15/08/96	25,61	3	3	.
64	4	.		1	1	0	0	.	.	.	.	10/11/51	24,61	3	2
27	3	1	14	1	0	0	0	ok	.	8	8	02/12/71	22,45	1	1
38	3	2	38	1	0	0	0	9/1-9/5	.	6	10	12/11/61	31,60	1	1
50	3	2	40	0	0	0	0	.	.	5	11	06/02/68	18,75	1	6
54	4	1	41	0	1	0	0	.	.	7	4	17/10/59	24,62	3	2
7	1	1	40	0	0	0	1	.	.	6	5	17/12/65	20,35	2	6
9	1	2	38	0	1	0	0	.	.	5	4	12/12/96	28,49	3	3
17	1	4	.	.	.	.	.	.	.		.	15/05/71	26,81	1	5
53	3	2	40	0	0	1	0	.	.	8	7	07/03/41	31,04	1	3
56	4	2	40	0	0	0	0	.	3,5	.	0	16/11/57	37,86	3	3
58	4	1	40	0	1	0	1	.	.	8	0	17/063/47	22,32	3	Y
14	1	1	38	0	0	0	1	.	.	7	12	04/05/61	19,12	4	2

Рис. 3-1. Проверка ошибок в наборе данных.

После введения данных, как описано в главе 2, набор данных следует проверить на наличие ошибок. Жирно обведённые цифры — ошибочно введённые данные. Например, код «41» в колонке «пол ребёнка» неверен, и в результате этого у больной 20 отсутствуют данные; оставшаяся часть данных была введена в неверные колонки. Другие (например, необычные данные в колонках срок беременности и масса) тоже похожи на ошибки, но эти пометки должны быть проверены, прежде чем будет принято какое-либо решение, так как они могут отразить подлинность выбросов. В этом случае срок беременности пациентки за номером 27 был 41 нед, а масса 11,19 записана неверно. Так как оказалось невозможным найти массу ребёнка, эти данные были введены (закодированы) как пропущенные.

[1] Простым и эффективным средством поиска ошибок ввода и выбросов является построение двухмерных диаграмм рассеяния для всех возможных пар признаков. Несмотря на то что уже при 10 признаках число таких графиков будет равно (10×9)/2=45, даже беглый просмотр этих графиков позволяет сразу же увидеть аномальные значения в виде точек, расположенных на значительном удалении от основного массива точек. Кроме того, полезно провести сортировку наблюдений по значениям одного столбца (признака) по убыванию и по возрастанию. В результате этого аномальные значения и ошибки ввода также будут сразу видны. Прим. переводчика.

4 Графическое представление данных

Первое, что вы захотите сделать после ввода данных в компьютер, — обобщить их таким образом, чтобы можно было «ощутить» их. Это можно сделать, создавая диаграммы, таблицы или статистическую сводку. Диаграммы — мощный инструмент передачи информации о данных для представления простых итоговых изображений, для обнаружения выбросов и тенденций до того, как будет проведён какой-либо запланированный анализ.

ОДНА ПЕРЕМЕННАЯ

Частотное распределение

Эмпирическое частотное распределение переменной связывает каждое возможное наблюдение, группу наблюдений (т.е. интервал значений) или категории с их наблюдаемой частотой появления. Если мы заменим каждую частоту относительной частотой (процент от общей частоты), то мы сможем сравнить распределения в двух и более группах индивидуумов.

Представление частотных распределений

Как только получены частоты (или относительные частоты) для категориальных или дискретных числовых данных, их можно наглядно представить.

Столбчатая или колончатая диаграмма

Для каждой категории чертят отдельный горизонтальный или вертикальный столбик, длина которого пропорциональна частоте данной категории. Столбики отделяются друг от друга небольшим пробелом для того, чтобы показать, какие это данные, категориальные или дискретные (рис. 4-1, а).

Круговая диаграмма

Круговая диаграмма делится на секции, каждой из которых отводится определённая категория, таким образом, чтобы площадь каждого сектора была пропорциональна частоте этой категории (рис. 4-1, б).

Бывает трудно отобразить непрерывные числовые данные: чтобы изобразить их графически, их нужно

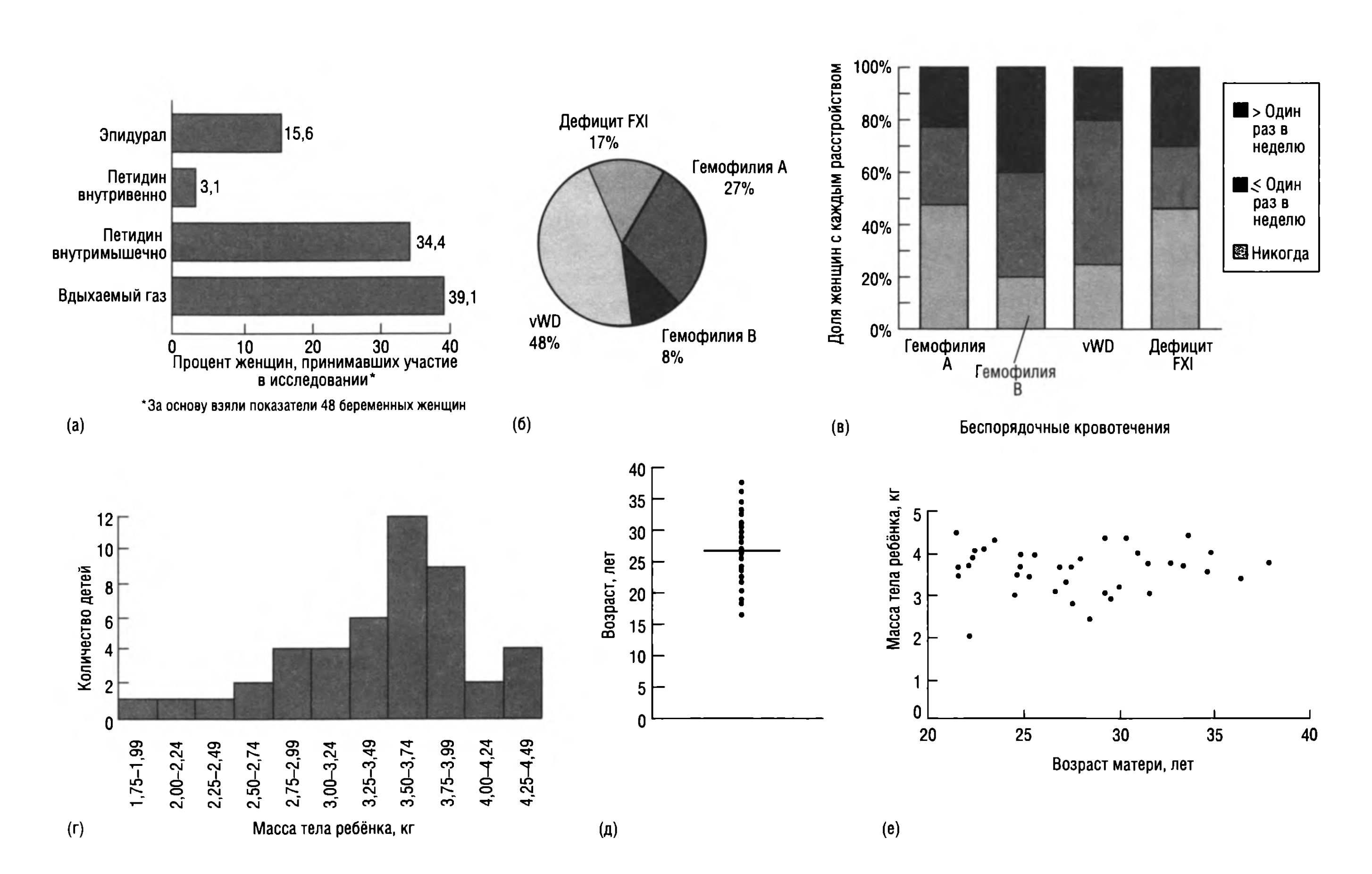


Рис. 4-1. Различные графические результаты, которые можно создать путём обобщения акушерских данных, полученных при исследовании женщин с беспорядочными кровотечениями (глава 2). Столбчатая диаграмма (а) показывает процентное отношение исследуемых женщин, которым необходимо облегчить боль во время родов от перечисленных вмешательств. Круговой график (б) показывает процентное отношение исследуемых женщин с беспорядочными кровотечениями. Сегментированная столбчатая диаграмма (в) показывает частоту, с которой женщины с беспорядочными различными кровотечениями испытывают кровотечения дёсен. Гистограмма (г) показывает массу тела ребёнка при рождении. Точечный график (д) показывает возраст матери во время рождения ребёнка со средним возрастом, который отмечен горизонтальной линией. Двухмерный график (е) показывает соотношение между возрастом матери во время родов (на горизонтальной, или x-оси) и массой тела ребёнка (на вертикальной, или y-оси).

сначала обобщить. Обычно используемые диаграммы включают следующее.

Гистограмма

Подобна круговой диаграмме, но здесь не должно быть пробелов между столбцами, так как данные непрерывны (рис. 4-1, г). Ширина каждого столбца гистограммы должна соответствовать интервалу значений данной переменной. Например, масса ребёнка (рис. 4-1, г) может быть от 1,75 до 1,99 кг, от 2 до 2,24 кг, от 4,25 до 4,49 кг. Площадь столбца пропорциональна частоте в данном интервале, поэтому если одна из групп охватывает более широкий интервал, чем другие, то основание столбца будет шире, а высота, соответственно, меньше. Обычно выбирают между пятью и 20 группами; интервал должен быть достаточно узким, чтобы отобразить структуру данных, но не должен быть настолько узким, чтобы они стали исходными данными (т.е. в каждом интервале по одному наблюдению). Гистограмма должна быть четко обозначена, чтобы было понятно, где находятся границы.

Точечный график

Каждое наблюдение изображено одной точкой на горизонтальной (или вертикальной) линии (рис. 4-1, д). Этот тип графика очень просто чертить, но только при небольшом объёме данных. Часто на диаграмме отображается обобщающая характеристика данных, такая, как среднее или медиана. Этот график можно использовать и для дискретных данных.

Беклометазон		Плацебо
3	1,0	04
665	1,1	39
53	1,2	99
9751	1,3	1135677999
955410	1,4	0148
987655	1,5	00338899
9531100	1,6	0001355
731	1,7	00114569
99843110	1,8	6
654400	1,9	01
6	2,0	
7	2,1	19
10	2,2	

Рис. 4-2. График «стебель и листья» показывает объём форсированного выдоха в литрах у детей, вдохнувших беклометазон или плацебо (глава 21).

График «стебель и листья»[1]

Это смесь диаграммы и таблицы; он похож на гистограмму и эффективен для отображения данных по увеличению порядка величины. Обычно чертят вертикальный стебель, который состоит из нескольких первых цифр данных, приведённых по порядку. Выходящие наружу от этого стебля листья — конечная цифра всех данных по порядку, которые написаны горизонтально (рис. 4-2) в порядке увеличения порядка расположения числа.

График *Box-plot*

Этот тип графиков часто называют «ящиком с усами». Это вертикальный или горизонтальный прямоугольник, где две параллельных стороны отвечают верхнему и нижнему квартилям данных. Линия, проведённая поперёк прямоугольника, отвечает значению среднего. «Усы», начинающиеся в конце прямоугольника, обычно показывают минимальные и максимальные значения, но иногда указывают и особые процентили, например, 5-й и 95-й процентили (глава 6, рис. 6-1). Здесь же могут быть обозначены и выбросы (аномально большие или малые значения).

Форма частотного распределения

Выбор наиболее подходящего статистического метода часто зависит от формы распределения. Распределение данных чаще всего унимодальное, т.е. имеющее одну «вершину». Иногда распределение бимодальное (две «вершины») или равномерное (каждая величина одинаково вероятна и нет «вершин»). При унимодальном распределении главная цель состоит в том, чтобы увидеть, где находится бо́льшая часть данных, относительно максимальных и минимальных значений. В частности, важно определить, каково распределение:

- симметричное — сосредоточенное вокруг средней точки, когда одна сторона — симметричное отражение другой (рис. 5-1);
- скошенное вправо (положительная асимметрия) — длинный правый «хвост» с одним или несколькими большими значениями. Такие данные являются весьма частыми в медицинском исследовании (рис. 5-2);
- скошенное влево (отрицательная асимметрия) — длинный левый «хвост» с одним или несколькими малыми значениями (рис. 4-1, г).

ДВЕ ПЕРЕМЕННЫЕ

Если одна переменная категориальная, тогда отдельные диаграммы, показывающие распределение второй переменной, должны быть начерчены для каждой категории. Другие графики, подходящие для таких данных, включают групповые или сегментные линии или графики с колонками (рис. 4-1, в).

Если обе переменные непрерывные или ординальные, то связь между ними можно изобразить при помощи двухмерной диаграммы рассеяния (скаттерплот) (рис. 4-1, е). Это двухмерный график, где оси переменных перпендикулярны друг другу. Одна переменная обычно называется *x* переменная и отображается на горизонтальной оси. Вторая переменная, известная как *y* переменная, наносится на вертикальную ось.

ИДЕНТИФИКАЦИЯ ВЫБРОСОВ ПРИ ИСПОЛЬЗОВАНИИ ГРАФИЧЕСКИХ МЕТОДОВ

Мы часто используем только одну переменную, отображающую данные, чтобы обнаружить выбросы. Например, длинный хвост на одной стороне гистограммы может указывать на удалённое, аномальное значение. Однако иногда выбросы могут стать очевидными только при рассмотрении соотношения между двумя переменными. Например, для женщины, рост которой 1,6 м, масса 55 кг не выглядит необычной, однако для женщины ростом 1,9 м такая масса будет необычно мала.

[1] В отечественных публикациях данный график незаслуженно не пользуется популярностью. Прим. переводчика.

5 Описание данных: «меры положения»

ОБОБЩЕНИЕ ДАННЫХ

Довольно трудно «прочувствовать» числовые измерения, до тех пор пока данные не будут обобщены содержательным образом. Диаграмма бывает полезна в качестве отправной точки. Также можно сжать информацию, представив величины, наиболее важные для характеристики данных. В частности, если знать, из чего состоит представленная величина или насколько широко рассеяны наблюдения, можно сформировать образ этих данных. Мера положения — общее понятие для числового выражения локализации (на числовой оси); которое описывает типичный результат измерения. Мы посвящаем эту главу мерам положения, самые распространённые из которых — среднее и медиана (табл. 5-1). Характеристики, которые отображают разброс или рассеяние наблюдений, мы включим в главу 6.

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Среднее арифметическое, которое очень часто называют просто среднее, для набора значений вычисляют следующим образом: складывают все значения и делят эту сумму на количество значений в этом наборе. Можно суммировать это буквальное выражение при помощи алгебраической формулы. Используя математическую систему обозначения, мы можем изобразить набор n наблюдений переменной x, как $x_1, x_2, x_3, .., x_n$. Например, x мог бы обозначать рост индивидуума (сантиметры), так чтобы x_1 обозначал рост первого индивидуума, а x_i — рост i индивидуума и т.д. Мы можем написать формулу для среднего арифметического наблюдений, пишется $\overline{x}$, а произносится «x с чертой»:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}.$$

Используя математическую систему обозначения, мы можем сократить это выражение:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

где Σ (греческая буква «сигма») означает «суммирование», а индексы внизу и вверху над этой буквой означают, что суммирование проводится от i=1 до i=n.

Это выражение часто сокращают ещё больше:

$$\overline{x} = \frac{\sum x_i}{n}, \text{ или } \overline{x} = \frac{\sum x}{n}.$$

МЕДИАНА

Если мы упорядочим данные по величине, начиная с самой маленькой величины и заканчивая самой большой, то медиана также будет характеристикой усреднения в упорядоченном наборе данных. Медиана делит ряд упорядоченных значений пополам, с равным числом этих значений как выше, так и ниже её (левее и правее медианы на числовой оси).

Вычислить медиану легко, если количество наблюдений n нечётное. Это будет наблюдение с номером $(n+1)/2$ в нашем упорядоченном наборе данных. Например, если n=11, то медиана — (11+1)/2=12/2=6, 6-е наблюдение в упорядоченном наборе данных. Если n чётное, тогда, строго говоря, медианы нет. Однако обычно мы вычисляем её как среднее арифметическое двух соседних средних наблюдений в упорядоченном наборе данных [т.е. наблюдений с номерами (n/2) и (n/2+1)]. Так, например, если n=20, то медиана — среднее арифметическое из наблюдений с номерами 20/2=10 и (20/2+1)=11 в упорядоченном наборе данных.

Медиана подобна среднему значению, если данные симметричны (рис. 5-1), меньше, чем среднее значение, если данные скошены вправо (рис. 5-2), и больше, чем среднее значение, если данные скошены влево.

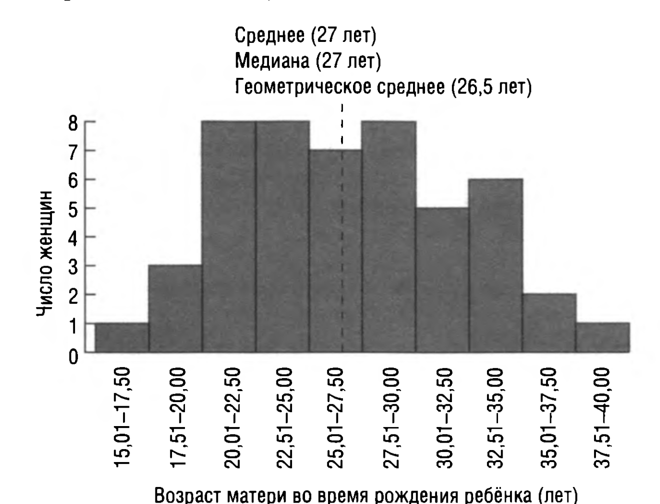


Рис. 5-1. Средняя, медиана и геометрическое среднее возраста женщин в исследовании, описанном в главе 2, во время рождения ребёнка. Распределение возраста довольно симметрично, поскольку все три «меры положения» дают близкие значения, обозначенные на графике пунктирной линией.

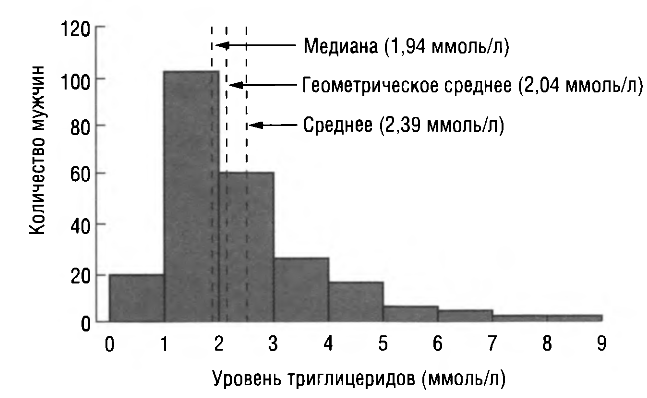


Рис. 5-2. Средняя, медиана и геометрическое среднее коцентрации триглицеридов в выборке из 232 мужчин с развившимся сердечным заболеванием (глава 19). Распределение концентрации триглицеридов скошено вправо, среднее арифметическое даёт более высокое значение «меры положения», чем медиана или геометрическое среднее.

МОДА

Мода — значение, которое встречается наиболее часто в наборе данных; если данные непрерывные, то мы обычно группируем их и вычисляем модальную группу. Некоторые наборы данных не имеют моды, потому что каждое значение встречается только один раз. Иногда можно встретить более одной моды; это происходит тогда, когда два значения или более встречаются одинаковое количество раз и частота встречаемости каждого из этих значений больше, чем таковые для любого другого значения. Мы редко используем моду как обобщающую характеристику.

СРЕДНЕЕ ГЕОМЕТРИЧЕСКОЕ

В случае если данные имеют несимметричное распределение, среднее арифметическое не будет обобщающим показателем такого распределения. Если данные скошены вправо, то можно создать распределение, которое будет более симметричным, если взять логарифм (по основанию 10 или по основанию *e*) каждого значения переменной в наборе данных. Среднее арифметическое значений этих логарифмов — характеристика распределения для преобразованных данных. Чтобы получить меру, которая будет иметь те же самые единицы измерения, как первоначальные наблюдения, мы должны осуществить обратное преобразование — потенцирование (т.е. взять антилогарифм) средней логарифмированных данных; мы называем такую величину среднее геометрическое. При условии, что распределение данных логарифма приблизительно симметричное, то среднее геометрическое подобно медиане и меньше, чем среднее необработанных данных (см. рис. 5-2).

ВЗВЕШЕННОЕ СРЕДНЕЕ

Мы используем взвешенное среднее в том случае, когда некоторые значения интересующей нас переменной x более важны, чем другие. Мы присоединяем массу w_i к каждому из значений x_i в выборке для того, чтобы учесть эту важность. Если значения x_1, x_2, x_3, .., x_n имеют соответствующие значения массы w_1, w_2, w_3, ... w_n, то взвешенное арифметическое среднее выглядит следующим образом:

$$\frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum w_ix_i}{\sum w_i}$$

Например, предположим, что мы заинтересованы в определении средней продолжительности пребывания госпитализированных больных в каком-либо районе и знаем средний реабилитационный период для больных в каждой больнице. Необходимо учитывать количество информации, в первом приближении принимая за вес каждого наблюдения такой показатель, как количество больных в больнице.

Взвешенное среднее и среднее арифметическое идентичны, если каждый вес равен единице.

Таблица 5-1. Преимущества и недостатки мер положения

Тип среднего	Преимущества	Недостатки
Среднее	Используются все значения набора данных. Определяется математически выполнимым алгебраическим выражением. Известно выборочное распределение (см. главу 6)	Искажается выбросами. Искажается асимметричными данными
Медиана	Не искажается выбросами. Не искажается асимметричными данными	Игнорирует большую часть информации. Не определяется алгебраически. Усложняется в выборочном распределении
Мода	Легко определяется для категориальных данных	Игнорирует большую часть информации. Не определяется алгебраически. Неизвестно выборочное распределение
Среднее геометриическое	До обратного преобразования имеет те же самые преимущества, что и среднее	Подходит, если логарифмическое преобразование образует симметричное распределение
Взвешенное среднее	Те же самые преимущества, что и у среднего. Приписывает соответствующий вес каждому наблюдению. Алгебраически определяется	Вес должен быть известен или оценён

6 Описание данных: «меры рассеяния»

ОБОБЩЕНИЕ ДАННЫХ

Если кратко изложить две меры непрерывной переменной, одна из которых показывает «средние» данные, а другая описывает «рассеяние» наблюдений, то данные удастся значительно сконцентрировать. Мы объяснили в главе 5, как выбрать соответствующую меру положения. В этой главе мы обсудим самые обычные способы описания меры рассеяния (разброса или вариабельности), сравнение которых приведено в табл. 6-1.

РАЗМАХ (ИНТЕРВАЛ ИЗМЕНЕНИЯ)

Размах — разность между максимальным и минимальным значениями переменной в наборе данных; вы найдёте эти две величины, на которые ссылаются вместо их разности. Обратите внимание, что этот размах вводит в заблуждение, если одно из значений — выброс (см. главу 3).

РАЗМАХ, ПОЛУЧЕННЫЙ ИЗ ПРОЦЕНТИЛЕЙ

Что такое процентили?

Предположим, что мы расположим наши данные упорядоченно, начиная с самой маленькой величины переменной x и заканчивая самой большой величиной. Величина x, до которой расположен 1% наблюдений, находящихся ниже x (99% наблюдений находятся выше значения x), называется первый процентиль. Величина x, до которой находится 2% наблюдений, называется второй процентиль и т.д. Величины x, которые делят упорядоченный набор значений на 10 равных групп, т.е. 10-й, 20-й, 30-й, .., 90-й процентили, называются децили. Величины x, которые делят упорядоченный набор значений на четыре равные группы, т.е. 25-й, 50-й и 75-й процентили, называются квартили. 50-й процентиль — медиана (см. главу 5).

Применение процентилей

Мы можем добиться такой формы описания рассеяния, на которую не повлияет выброс (аномальное значение), при этом исключая экстремальные величины в наборе данных и определяя размах остающихся наблюдений. Межквартильный размах — разница между первым и третьим квартилем, т.е. между 25-м и 75-м процентилями (рис. 6-1). В него входят центральные 50% наблюдений в упорядоченном наборе, где 25% наблюдений находятся ниже центральной точки и 25% — выше. Интердецильный размах содержит в себе центральные 80% наблюдений, т.е. те наблюдения, которые располагаются между 10-м и 90-м процентилями. Часто используют размах, который содержит 95% наблюдений, т.е. он исключает 2,5% наблюдений снизу и 2,5% сверху (см. рис. 6-1). Можно применить этот интервал, осуществляя диагностику болезни. В этом случае он называется референтный интервал, референтный размах или нормальный размах.

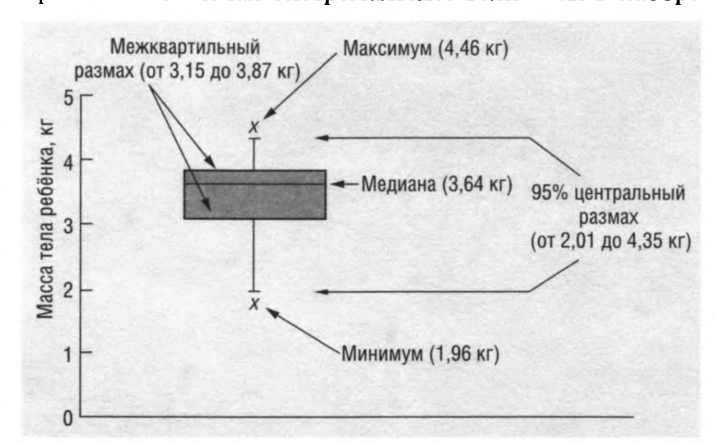


Рис. 6-1. График «ящик с усами». Масса ребёнка при рождении (см. главу 2). На этом рисунке отображена медиана, межквартильный размах, размах, который содержит центральные 95% наблюдений и максимальную и минимальную величины.

ДИСПЕРСИЯ[1]

Один из способов измерения рассеяния данных заключается в том, чтобы определить степень отклонения каждого наблюдения от средней арифметической. Очевидно, что чем больше отклонение, тем больше изменчивость, вариабельность наблюдений. Однако мы не можем использовать среднее этих отклонений как меру рассеяния, потому что положительные отклонения компенсируют отрицательные отклонения (их сумма тождественно равна нулю). Для того чтобы решить эту проблему, мы возводим в квадрат каждое отклонение и находим среднее возведённых в квадрат отклонений (рис. 6-2); эта величина называется вариацией, или дисперсией. Возьмем, например, n наблюдений, $x_1, x_2, x_3, .., x_n$, средняя которых равняется $\overline{x} = (\sum x_i)/n$, мы вычисляем дисперсию, обычно обозначаемую как s^2, этих наблюдений следующим образом:

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}.$$

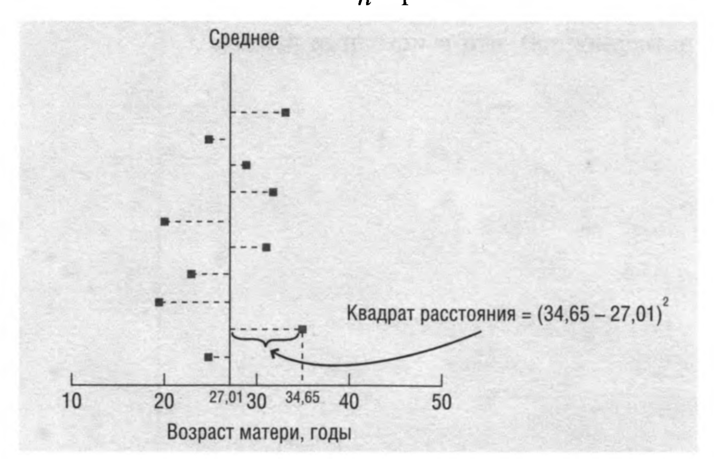


Рис. 6-2. Диаграмма показывает рассеяние выбранных значений возраста матери во время рождения ребёнка (см. главу 2) вокруг средней величины. Вариация вычисляется путём суммирования разностей от каждой точки до среднего значения, возведённых в квадрат, и делением этой суммы на (n–1).

[1] Название этой статистики происходит от латинского *dispersus* — рассеянный, рассыпанный. Прим. переводчика.

Таблица 6-1. Преимущества и недостатки мер рассеяния

Мера рассеяния	Преимущества	Недостатки
Размах	Легко определить	Использует только два наблюдения. Искажается выбросами. Имеет тенденцию к увеличению при росте объёма выборки
Размахи, основанные на процентилях	Не подвержены влиянию выбросов. Не зависят от размера выборки. Подходят для асимметричных распределений данных	Грубый расчёт. Невозможно рассчитать для маленьких выборок. Может использоваться только при двух наблюдениях. Не имеет алгебраического выражения для вычисления
Дисперсия	Использует каждое наблюдение. Имеет алгебраическое выражение для вычисления	Единица измерения — квадрат исходных данных. Восприимчива к выбросам. Не подходит для асимметричных распределений данных
Стандартное отклонение	Те же самые преимущества, что и у вариации. Единицы измерения те же, что и у исходных данных. Легко интерпретируема	Восприимчива к выбросам. Не подходит для асимметричных распределений данных

Мы видим, что это не то же, что среднее арифметическое возведённых в квадрат отклонений, потому что мы разделили на (n–1) вместо n. Причина этого в том, что мы почти всегда полагаемся на выборочные данные в наших исследованиях. Теоретически можно показать, что мы получим более точную дисперсию, если разделим не на n, а на (n–1).

Единица измерения (размерность) вариации — квадрат единиц измерения первоначальных наблюдений, например: если вариация измеряется в килограммах, то единицей измерения вариации будет «кг^2».

СТАНДАРТНОЕ ОТКЛОНЕНИЕ

Стандартное (среднее квадратичное) отклонение — положительный квадратный корень из дисперсии. На примере n наблюдений это выглядит так:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}.$$

Мы можем размышлять о стандартном отклонении как о своего рода среднем отклонении наблюдений от среднего. Его вычисляют в тех же самых единицах (размерностях), что и исходные данные.

Если разделить стандартное отклонение на среднее арифметическое и выразить этот показатель в процентах, получится коэффициент вариации. Это мера рассеяния, которая не зависит от единиц измерения (безразмерная), но имеет некоторые теоретические неудобства, поэтому статистики её не всегда одобряют.

ИЗМЕНЕНИЯ МЕЖСУБЪЕКТНЫЕ И ВНУТРИСУБЪЕКТНЫЕ

Если провести повторные измерения непрерывной переменной у одного и того же пациента, то можно увидеть, какие изменения происходят в ответах этого индивидуума (внутрисубъектные изменения). Причиной этого могут быть не всегда точные ответы данного индивидуума и/или ошибка измерения. Однако если проводить измерения у одного пациента, то вариации обычно меньше, чем если проводить единичные измерения на каждом индивидууме в группе (межсубъектные изменения). Например, вместимость лёгкого 17-летнего юноши между 3,6 и 3,87 л при повторных измерениях (не менее 10 раз); если же проводить однократные измерения на 10 юношах того же возраста, то объём будет между 2,98 и 4,33 л. Эти концепции важны в плане исследования.



Теоретические распределения: нормальное распределение

В главе 4 мы показали, как создать эмпирическое распределение частоты исследуемых данных. Оно контрастирует с теоретическим распределением вероятности, которое можно описать при помощи математической модели. Когда эмпирическое распределение аппроксимирует некоторое распределение вероятности, то можно применить теоретические знания об этом распределении для того, чтобы ответить на вопросы относительно данных. Часто это требуется для оценки вероятностей.

ПОНИМАНИЕ ВЕРОЯТНОСТИ[1]

Вероятность измеряет неопределённость. Она находится в самом центре статистической теории. Вероятность измеряет возможность появления данного события. Это положительное число, которое находится в интервале между 0 и 1. Если она равна нулю, то никакого события и быть не может. Если она равна единице, тогда событие должно обязательно произойти. Вероятность дополнительного события (события не происходящего) равна единице минус вероятность появления события. Мы рассмотрим условную вероятность как вероятность события с учётом того, что другое событие произошло, в главе 45.

Можно вычислить вероятность, используя различные подходы.

- Субъективная — индивидуальная степень уверенности, что данное событие произойдёт (например, случится конец света в 2050 г.).
- Частотная — выражающая соотношение количества событий, которые могли бы произойти, если бы мы повторяли эксперимент огромное количество раз (например, если бы мы бросали монету 1000 раз, сколько бы раз выпал «орёл»).
- Априорная — требующая знания теоретической модели, называемой распределением вероятности, которая отображает вероятности всех возможных результатов «эксперимента». Например, генетическая теория позволяет нам отобразить вероятность распределения цвета глаз у ребёнка при рождении, если у женщины голубые глаза, а у мужчины карие, первоначально определяя весь возможный генотип цвета глаз у ребёнка и их вероятности.

ПРАВИЛА ВЕРОЯТНОСТИ

Мы можем применять правила вероятности для того, чтобы складывать и умножать вероятности.

- Правило сложения: если два события, А и В, взаимоисключающие, несовместимые (т.е. каждое событие исключает другое), то вероятность того, что произойдёт то или иное событие, равна сумме их вероятностей. Prob(A или B)=Prob(A)+Prob(B). Например, вероятности того, что у взрослого больного в обычной зубоврачебной практике присутствуют все зубы, некоторые зубы отсутствуют или нет зубов, равны 0,67, 0,24 и 0,09 соответственно. Тогда вероятность того, что у больного есть несколько зубов, равна 0,67+0,24=0,91.
- Правило умножения: если два события, А и В, независимы (т.е. возникновение одного события не влияет на возможность появления другого), то вероятность того, что оба события произойдут, равна произведению вероятности каждого: Prob(A и B) = Prob(A)×Prob(B). Например, если двое, не имеющих отношения друг к другу, больных ожидают в кабинете хирургической стоматологии, то вероятность того, что у обоих больных присутствуют все зубы, равна 0,67×0,67=0,45.

РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ: ТЕОРИЯ

Случайная переменная — это величина, которая может принимать любое из набора взаимоисключающих значений с определённой вероятностью. Распределение вероятности показывает вероятности всех возможных значений случайной переменной. Это теоретическое распределение, которое выражено математически и имеет среднее и дисперсию, являющиеся аналогами среднего и дисперсии в эмпирическом распределении. Каждое распределение вероятности обусловлено параметрами, обобщающими величинами (например, среднее, дисперсия), характеризующими данное распределение (т.е. знание их позволит подробно описать распределение). При помощи соответствующей статистики можно провести оценку этих параметров в выборке. В зависимости от того, относится ли случайная переменная к дискретным или непрерывным данным, распределение вероятности может быть либо дискретное, либо непрерывное.

- Дискретное (например, биноминальное, распределение Пуассона). Можно получить вероятности, соответствующие каждому возможному значению случайной переменной. Сумма всех таких вероятностей равна единице.
- Непрерывное (например, нормальное, Хи-квадрат, t и F). Можно получить вероятность случайной переменной x, только принимающей значения в определённых интервалах (потому что существует бесконечное множество значений x). Если горизонтальная ось изображает значения x, то можно начертить кривую из уравнения распределения (функция плотности распределения вероятности); она имеет сходство с эмпирическим относительным частотным распределением (см. главу 4). Общая площадь под кривой равна единице; эта площадь отражает вероятность всех возможных событий. Вероятность того, что x находится между двумя пограничными значениями,

[1] Понятие вероятности не самое простое для глубокого понимания и осмысления. Поэтому не случайно оно постоянно уточнялось и дополнялось на протяжении нескольких столетий. В конце XX в., в частности, было установлено, что возможно оценить вероятность появления далеко не всех событий, поэтому в отношении таких событий был введен термин «неопределённые события», т.е. такие события, для которых вероятность неопределима. Прим. переводчика.

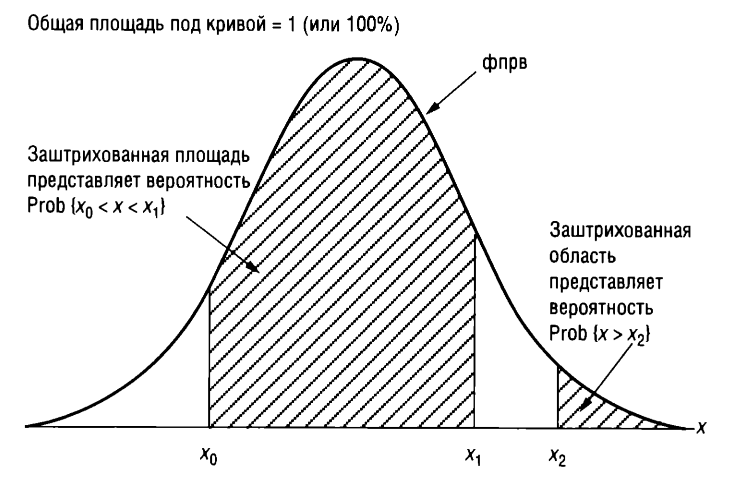


Рис. 7-1. Функция плотности распределения вероятности, *x*.

равна площади под кривой между этими значениями (рис. 7-1). Для удобства была создана таблица (приложение A) для того, чтобы было возможно оценить вероятность для обычно используемых непрерывных распределений вероятности. В особенности они важны при построении доверительных интервалов (см. главу 11) и проверке гипотез (см. главу 17).

НОРМАЛЬНОЕ (ГАУССОВСКОЕ) РАСПРЕДЕЛЕНИЕ

Одно из самых важных распределений в статистике — нормальное распределение. Его функция плотности распределения вероятности (рис. 7-2):

- полностью определяется двумя параметрами, среднее (μ) и дисперсия (σ^2);
- колоколообразна (унимодальна);
- симметрична относительно среднего;
- сдвигается вправо, если среднее увеличивается, и влево, если среднее уменьшается (при постоянной дисперсии);
- сплющивается, если дисперсия увеличивается, но становится более остроконечной, если дисперсия уменьшается (для постоянного среднего).

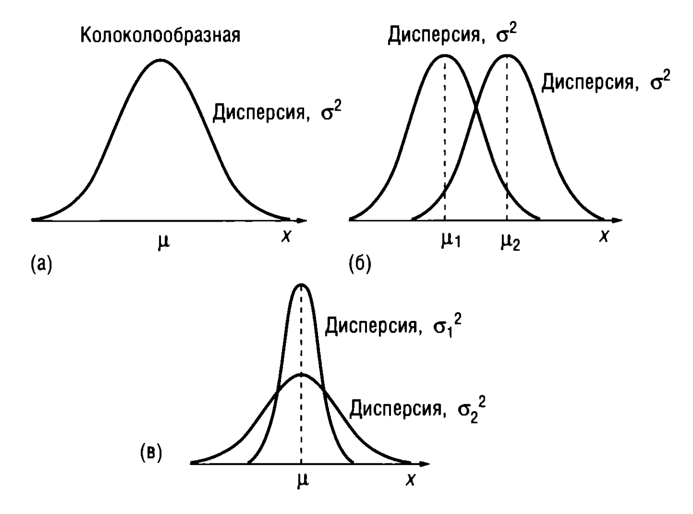


Рис. 7-2. Функция плотности распределения вероятности нормального распределения переменной *x*: (а) симметричное относительно среднего μ, дисперсия σ_2; (б) результат изменения среднего ($\mu_2 > \mu_1$); (в) результат изменения дисперсии ($\sigma_1^2 < \sigma_2^2$).

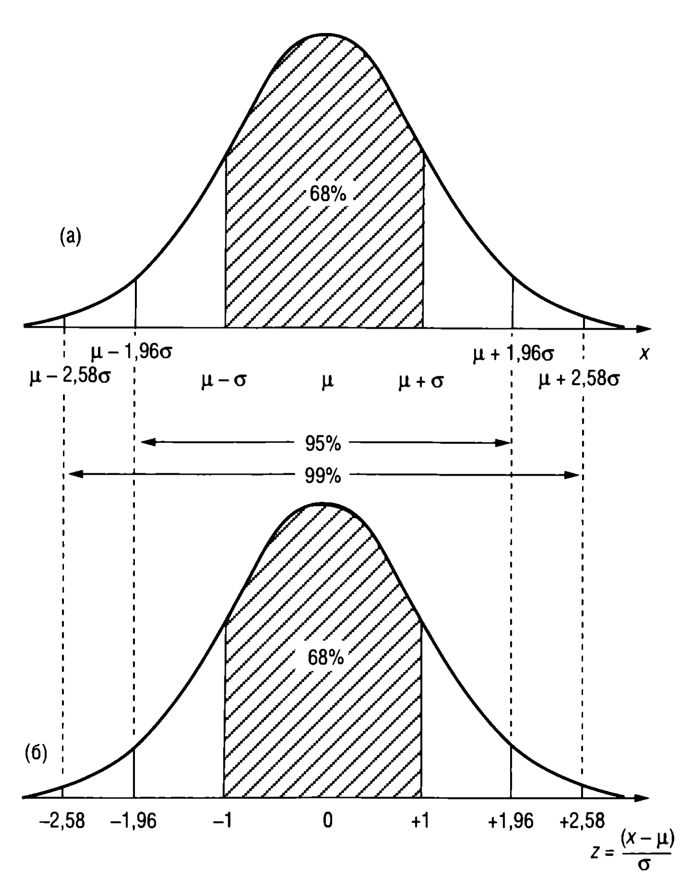


Рис. 7-3. Площади (проценты общей вероятности) под кривой для (а) нормального распределения *x*, со средним μ и дисперсией σ_2 и (б) стандартного нормального распределения *z*.

Дополнительные свойства.

- Среднее и медиана нормального распределения равны.
- Вероятность (рис. 7-3, а) того, что нормально распределённая случайная переменная *x*, со средним μ и стандартным отклонением σ, находящаяся между:
  - ✧ ($\mu-\sigma$) и ($\mu+\sigma$), равна 0,68;
  - ✧ ($\mu-1{,}96\sigma$) и ($\mu+1{,}96\sigma$), равна 0,95;
  - ✧ ($\mu-2{,}58\sigma$) и ($\mu+2{,}58\sigma$), равна 0,99.

Эти интервалы можно использовать для определения референтных интервалов (главы 6 и 35).

В главе 32 мы покажем, как оценить нормальность.

СТАНДАРТНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Существует бесконечно много (семейство) нормальных распределений в зависимости от значений μ и σ. Стандартное нормальное распределение — особое нормальное распределение, вероятности которого приведены в таблицах A1, A4 (приложение A).

- Стандартное нормальное распределение имеет среднее, равное 0, и дисперсию, равную 1.
- Если случайная переменная *x* имеет нормальное распределение со средним μ и дисперсией σ^2, тогда стандартизованное нормальное отклонение (СНО) будет случайной переменной, которая имеет стандартное нормальное распределение.

$$z = \frac{x-\mu}{\sigma}.$$

8 Теоретические распределения: другие распределения

НЕСКОЛЬКО СЛОВ ДЛЯ ПОДДЕРЖКИ ДУХА

Не волнуйтесь, если теория, лежащая в основе распределения вероятности, вам покажется сложной. Наш опыт показывает: единственное, что вы хотели бы знать, когда и как применять эти распределения, поэтому мы изложили основы и опустили уравнения, которые характеризуют распределения вероятности. И вы поймете, что вам нужно освоить основные понятия, терминологию и, возможно, знать, как работать с таблицами.

НЕПРЕРЫВНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Эти распространения основаны на непрерывных случайных переменных. Часто это не непосредственно измеряемая переменная, которая отвечает такому распределению, а параметр, статистика, полученная из этой переменной. Общая площадь под кривой функции плотности распределения вероятности есть сумма вероятностей всех возможных значений и равна 1 (см. главу 7). Мы рассмотрели нормальное распределение в главе 7; в этой главе мы опишем другие наиболее известные распределения.

t-распределение

См. приложение А, табл. А2, рис. 8-1.

- Получено Вильямом Госсетом, который публиковался под псевдонимом Студент (Student), поэтому его часто называют *t*-распределением Стьюдента.
- Параметры, которые характеризуют *t*-распределение, — это степени свободы (*df*), так мы сможем начертить функцию плотности распределения вероятности только в том случае, если мы будем знать уравнение *t*-распределения и степени свободы. В главе 11 мы рассмотрим степени свободы; обратите внимание, что они часто выражаются через объём выборки.
- Форма подобна форме для стандартизованного нормального распределения, но более приплюснута и с более длинными хвостами. Форма приближается к нормальной кривой, по мере того как увеличиваются степени свободы.

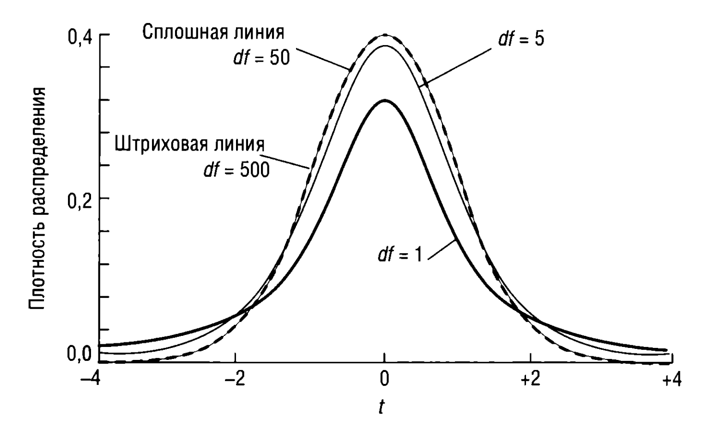


Рис. 8-1. *t*-распределение со степенями свободы (*df*)=1; 5; 50 и 500.

- В частности, его применяют для вычисления доверительных интервалов и исследования гипотез с одной или двумя средними.

Хи-квадрат

Хи-квадрат, (χ^2) или распределение Пирсона (приложение А, табл. А3; рис. 8-2):

- скошено вправо и принимает только положительные значения;
- характеризуется степенями свободы;
- его форма зависит от числа степеней свободы — становится более симметричной и приближается к нормальной с их ростом;
- особенно часто используется для анализа категориальных данных (см. главы 23—25).

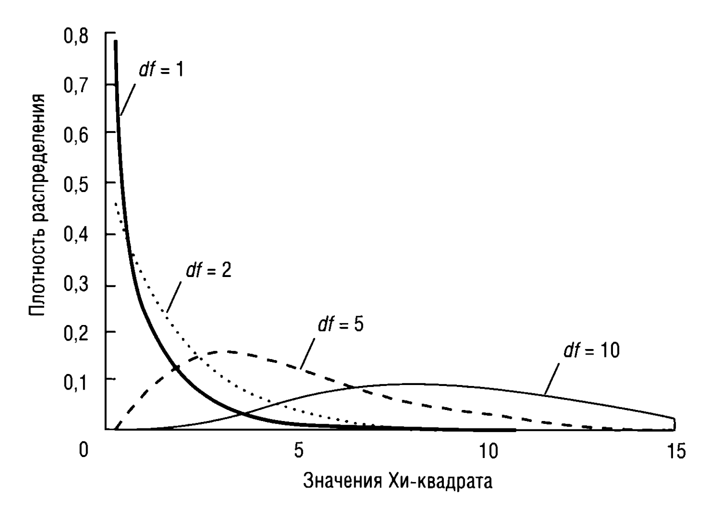


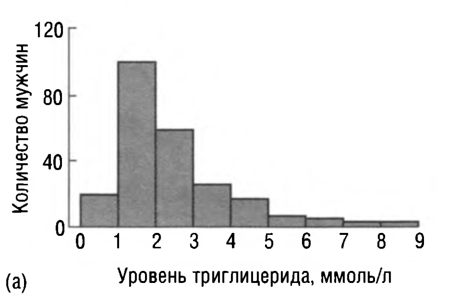
Рис. 8-2. Хи-квадрат распределения Пирсона со степенями свободы (*df*)=1; 2; 5 и 10.

F-распределение (см. табл. А5 в приложении А)

- Скошено вправо.
- Определяется как отношение. Распределение отношения двух оценок дисперсий, вычисленных для нормально распределённых данных, аппроксимируются *F*-распределением.
- Два параметра, которые характеризуют его, — степени свободы (см. главу 11) числителя и знаменателя отношения.
- *F*-распределение особенно полезно для сравнения двух дисперсий (см. главу 18) и более чем двух средних при использовании дисперсионного анализа (ANOVA).

Логнормальное распределение

- Распределение вероятности случайной переменной, логарифм которого (по основанию 10 или *e* — основание натурального логарифма) имеет нормальное распределение.
- Сильно скошено вправо (рис. 8-3, а).
- Если взять логарифмы исходных данных, которые скошены вправо, мы создадим эмпирическое распределение, которое почти нормальное (рис. 8-3, б),



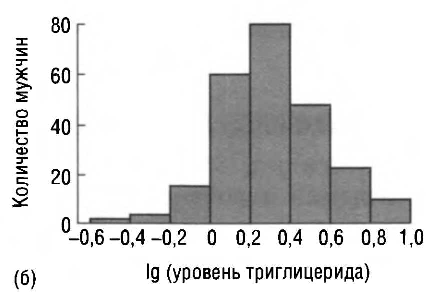


Рис. 8-3. (а) Логнормальное распределение уровня триглицерида у 232 мужчин с заболеваниями сердца (см. главу 19). (б) Почти нормальное распределение lg (уровень триглицерида).

и тогда данные соответствуют приближённо логнормальному распределению.

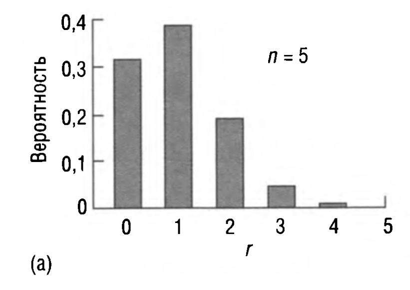
- Многие переменные в медицине имеют логнормальное распределение. Можно использовать свойства нормального распределения для того, чтобы сделать выводы относительно этих переменных после логарифмического преобразования данных.
- Если набор данных имеет логнормальное распределение, то используют среднее геометрическое как обобщающий показатель положения.

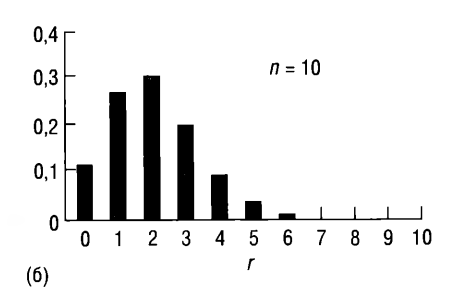
ДИСКРЕТНЫЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Дискретной называют случайную переменную, которая имеет такое распределение вероятности. Сумма вероятностей всех возможных взаимоисключающих событий равна 1.

Биноминальное распределение

- Предположим, в данной ситуации существует только два результата: «успех» и «неудача». Например, нас интересует, забеременеет или нет женщина при экстракорпоральном оплодотворении (ЭКО). Если посмотреть на n=100, не имеющих отношения друг к другу, женщин, подвергающихся ЭКО (каждая с той же самой вероятностью беременности), то биноминальная случайная переменная — наблюдаемое количество зачатий. Часто это понятие объясняется в терминах n независимых повторных испытаний (например, 100 раз подбросить монету), при которых результатом будет либо успех (например, орёл), либо неудача.
- Два параметра, которые описывают биноминальное распределение, — n, количество индивидуумов в выборке (или повторения испытания), и π, точная вероятность успеха для каждого индивидуума (или при каждом испытании).
- Среднее (значение для случайной переменной, которое мы ожидаем, если мы осматриваем n индивидуумов или повторим испытание n раз) равно $n\times\pi$. Дисперсия — $n\times\pi\times(1-\pi)$.
- Когда n невелико, распределение будет скошено вправо, если π <0,5, и влево, если π >0,5. Распределение становится более симметричным, по мере того как объём выборки увеличивается (рис. 8-4), и приблизится к нормальному распределению в том случае, если $n\times\pi$ и $n\times(1-\pi)$ больше 5.
- Можно использовать свойства биноминального распределения для того, чтобы сделать выводы относительно пропорций. Особенно часто используют аппроксимацию биномиального распределения с помощью нормального распределения при анализе пропорций.





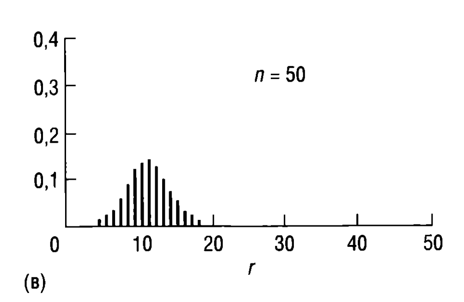


Рис. 8-4. Биноминальное распределение показывает количество успехов r, когда вероятность успеха π=0,2 для объёма выборки (а) n=5, (б) n=10 и (в) n=50 (N.B. в главе 23, наблюдаемая серораспространённость вируса герпеса человека была p=0,187≈0,2, и объём выборки был 271: для пропорции было использовано приближение нормальным распределением).

Распределение Пуассона

- Пуассоновская случайная переменная — число событий, которые происходят независимо и случайно во времени или пространстве с постоянной средней интенсивностью μ. Например, количество госпитализаций в день типично отвечает распределению Пуассона. Знание распределения Пуассона используют для того, чтобы вычислить вероятность конкретного количества госпитализаций в любой отдельный день.
- Параметр, которым описывают распределение Пуассона, — среднее, т.е. средняя интенсивность, μ.
- Среднее равняется дисперсии в распределении Пуассона.
- Если среднее ближе к минимальному, то распределение будет скошено вправо и становится более симметричным по мере того, как среднее будет увеличиваться, оно приближается, по форме, к нормальному распределению.

9 Преобразования

ЗАЧЕМ ПРЕОБРАЗОВЫВАТЬ?

В нашем исследовании наблюдения могут не подчиняться требованиям предполагаемого статистического анализа.

- Переменная может иметь ненормальное распределение, требование, обязательное для множества различных видов анализа.
- Рассеяние наблюдений в каждой из групп ряда может быть различным (равенство дисперсий — необходимое предположение относительно параметра при сравнении средних, использующем t-критерий и дисперсионный анализ, см. главы 21—22).
- Две переменные не могут быть линейно зависимы (линейность — требование во многих видах регрессионного анализа, см. главы 27—31).
- Очень часто полезно преобразовывать данные для того, чтобы удовлетворить требованиям, которые лежат в основе предлагаемых статистических методов[1].

КАК МЫ ПРЕОБРАЗОВЫВАЕМ?

Мы превращаем исходные данные в преобразованные, используя одно и то же математическое преобразование для каждого наблюдения. Предположим, что есть n наблюдения ($y_1, y_2, .., y_n$) с переменной y, и мы считаем, что подходит логарифмическое преобразование. Берём логарифм каждого наблюдения для того, чтобы образовать ($\log y_1, \log y_2, .., \log y_n$). Если назвать преобразованную переменную z, тогда $z = \log y_i$ для каждой i ($i = 1, 2, .., n$), и наше преобразование данных можно записать ($z_1, z_2, .., z_n$).

Проверяем, достигло ли это преобразование своей цели при создании набора данных, который удовлетворяет предположениям запланированного статистического анализа, и переходим к анализу преобразования данных ($z_1, z_2, .., z_n$). Часто делают обратные преобразования обобщающих мер (таких, как среднее) до первоначальной размерности; те выводы, которые мы сделали из гипотез по тестам на преобразованных данных, применимы для исходных данных[2].

ТИПИЧНЫЕ ПРЕОБРАЗОВАНИЯ

Логарифмическое преобразование, $z = \log y$

При логарифмическом преобразовании данных можно выбрать, взять логарифм или по основанию 10 ($\log_{10}y = \lg y$, «обычный» логарифм), или по основанию e ($\log_e y = \ln y$, «натуральный», или неперовский логарифм), но они должны быть одинаковы для отдельной переменной в наборе данных. Обратите внимание, что нельзя брать логарифм отрицательного числа или нуля. Обратное преобразование логарифма (потенцирование) называется антилогарифмом; антилогарифм неперовского логарифма есть экспонента e.

- Если распределение y скошено вправо, преобразование $z = \log y$ часто даёт в результате приближённо нормальное распределение (рис. 9-1, а). Тогда y имеет логнормальное распределение.
- Если существует показательное соотношение между y и другой переменной x, такое, что конец кривой загибается вверх, в то время как y (по вертикальной оси) наносится соответственно x (по горизонтальной оси), то соотношение между $z = \log y$ и x приблизительно линейное (рис. 9-1, б).
- Предположим, что есть различные группы наблюдений, каждая включает в себя измерения непре-

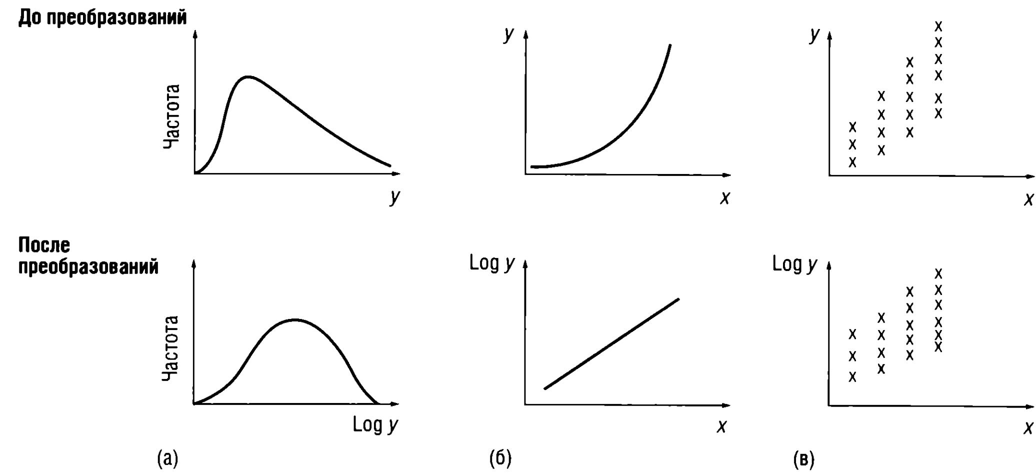


Рис. 9-1. Свойства логарифмического преобразования: (а) нормализация, (б) линеаризация, (в) выравнивание дисперсий.

[1] При этом важно помнить, что результат применения статистического критерия к преобразованной переменной нельзя априори переносить на исходную переменную. Далее, преобразование может изменить и размерность числовой переменной, что затрудняет смысловую интерпретацию новой переменной. Например, систолическое давление имеет размерность [мм рт.ст.]. После возведения в квадрат этой переменной размерность станет [мм рт.ст.]^2. Каков физический смысл этой новой переменной? Прим. переводчика.

[2] Такое согласие наблюдается далеко не всегда, и потому оно требует дополнительной проверки. Прим. переводчика.

рывной переменной y. Можно обнаружить, что группы, которые имеют более высокие значения y, имеют также и большие вариации. В частности, если коэффициент вариации (стандартное отклонение, делённое на среднее) постоянен для всех групп, то преобразование логарифма $z = \log y$ создаёт группы с равными дисперсиями (рис. 9-1, в).

В медицине часто применяют логарифмическое преобразование из-за логической интерпретации и потому, что многие переменные имеют скошенное вправо распределение.

Преобразование квадратного корня, $z = \sqrt{y}$

Это преобразование имеет свойства такие же, как и лог-преобразование, хотя результаты, после того как они были преобразованы обратно, объяснить сложнее. В дополнение к способностям нормализации и линеаризации эффективно использовать преобразование, стабилизирующее дисперсию, если дисперсия возрастает при увеличении значений y, т.е. тогда дисперсия, делённая на среднее, постоянна. Мы применяем преобразование квадратного корня, если y — количество редких явлений, встречающихся во времени и пространстве, т.е. это пуассоновская переменная. Помните, что мы не можем извлечь квадратный корень из отрицательного числа.

Обратное преобразование, $z = 1/y$

Часто применяют обратное преобразование к периодам жизни (выживаемости), если не используют специальные методы для анализа выживаемости. Обратное преобразование имеет свойства такие же, как и лог-преобразование. В дополнение к способностям нормализации и линеаризации оно гораздо эффективнее для стабилизации дисперсии, чем логарифмическое преобразование, если дисперсия очень заметно увеличивается при увеличении значений y, т.е. для стабилизации частного от деления дисперсии на среднее. Заметьте, что нельзя делить на нуль.

Квадратичное преобразование, $z = y^2$

Квадратичное преобразование достигает результата, обратного лог-преобразованию.

- Если y скошено влево, то распределение $z = y^2$ часто приближённо нормальное (рис. 9-2, а).
- Если соотношение между двумя переменными x и y такое, что, когда мы наносим y против x, кривая загибается вниз, тогда соотношение между $z = y^2$ и x почти линейное (рис. 9-2, б).
- Если дисперсия непрерывной переменной y уменьшается, по мере того как значение y увеличивается, тогда квадратное преобразование $z = y^2$ стабилизирует дисперсию (рис. 9-2, в).

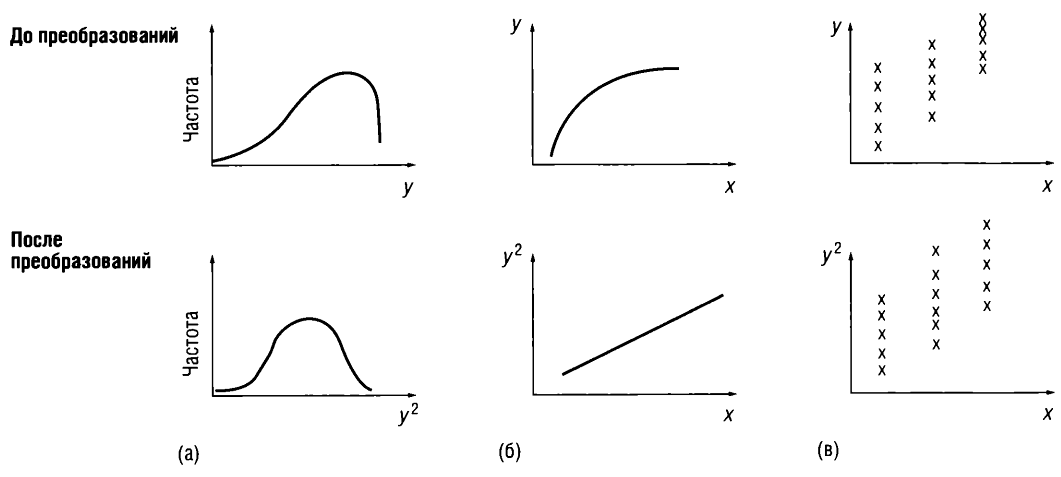


Рис. 9-2. Свойства квадратичного преобразования: (а) нормализация, (б) линеаризация, (в) выравнивание дисперсий.

Логит-преобразование $z = \ln \frac{p}{1-p}$

Это преобразование, которое наиболее часто применяют к каждой пропорции p, в наборе пропорций. Невозможно применять логистическое преобразование в том случае, если $p=0$ либо $p=1$, потому что соответствующие значения логита — $-\infty$ и $+\infty$. Единственное решение — взять p как $1/(2n)$ вместо 0 и как $\{1-1/(2n)\}$ вместо 1.

Оно линеаризирует сигмовидную кривую (рис. 9-3).

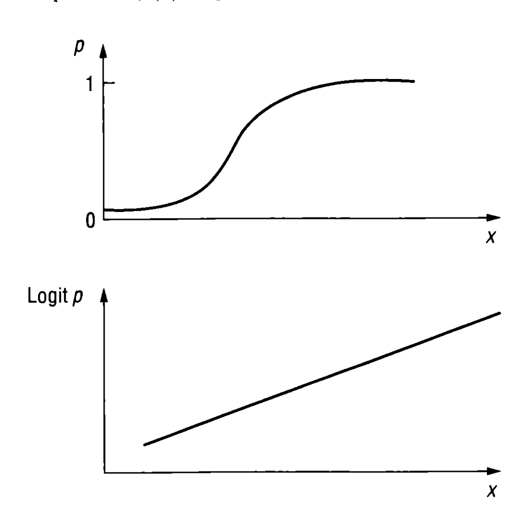


Рис. 9-3. Свойство логистического преобразования сигмовидной кривой.

10 Выборка и выборочное распределение

ЗАЧЕМ МЫ ДЕЛАЕМ ВЫБОРКУ?

В статистике популяция представляет целую группу индивидуумов, которые нас интересуют. Вообще, изучать целую популяцию довольно дорого и трудоёмко, а в некоторых случаях просто невозможно, так как популяция может быть гипотетической (например, пациенты, которые будут проходить лечение в будущем), поэтому собирают данные по выборке индивидуумов, как предполагают, представителей этой популяции, и используют их для того, чтобы сделать выводы (т.е. делать заключения) относительно этой популяции.

Когда берут выборку из популяции, имеют в виду, что информация в выборке не может полностью отражать то, что истинно в этой популяции. Возможна ошибка, обусловленная выборкой, так как изучалась только часть популяции. В этой главе мы покажем, как использовать теоретическое распределение вероятности для определения величины этой ошибки.

ПОЛУЧЕНИЕ РЕПРЕЗЕНТАТИВНОЙ (ПРЕДСТАВИТЕЛЬНОЙ) ВЫБОРКИ

В идеале мы стремимся к случайной (рандомизированной) выборке. Составляют список всех индивидуумов в популяции (структура выборки) и из этого списка случайно отбирают индивидуумов, т.е. каждая возможная выборка данного объёма имеет равную вероятность быть выбранной из популяции. Иногда возникают трудности при составлении этого списка или материальные ограничения, и в этом случае берут приемлемую выборку. Например, при исследовании пациентов с особым клиническим заболеванием можно выбрать одну больницу или несколько, исследовать несколько или всех пациентов с этим заболеванием в этой (или каждой) больнице. Иногда можно использовать неслучайные схемы, такие, как квотированная выборка или систематическая выборка. Хотя статистические тесты (критерии), описанные в этой книге, предполагают, что индивидуумы отобраны случайно, такие методы вообще приемлемы, пока выборка включает представителей этой популяции.

ТОЧЕЧНЫЕ ОЦЕНКИ

Мы часто заинтересованы в оценке параметра в популяции, среднего или пропорции (доли). Параметры обычно обозначают символами греческого алфавита. Например: обычно мы обозначаем среднее популяции как μ, а стандартное отклонение популяции как σ[1].

Мы оцениваем значение параметра, используя данные, собранные в выборке. Эта оценка — точечная оценка генерального параметра (т.е. она принимает только одно значение) в отличие от интервальной оценки, которая имеет интервал значений. Точечную оценку описывает выборочная статистика.

[1] В статистике принято обозначать популяционные параметры (генеральные) буквами греческого алфавита, а выборочные — соответствующими им буквами латинского алфавита, например, μ и *m*, σ и *s* и т.д. Прим. переводчика.

ВЫБОРОЧНАЯ ДИСПЕРСИЯ

Если повторить извлечение выборок того же самого объёма из популяции, маловероятно, что оценки параметра популяции будут точно такими же в каждой выборке. Однако все оценки должны быть близки к истинному значению параметра (генеральному параметру) в популяции и подобны друг другу. Определяя величину вариабельности этих оценок, мы поймём, насколько они точны, и таким образом сможем оценить ошибку, обусловленную выборкой. Обычно берут только одну выборку из популяции. Однако можно использовать знания о теоретическом распределении выборочных оценок для того, чтобы сделать выводы относительно генерального параметра популяции.

ВЫБОРОЧНОЕ РАСПРЕДЕЛЕНИЕ СРЕДНЕГО

Предположим, что мы заинтересованы в оценке среднего популяции; можно брать много повторных выборок объёма n из популяции и оценить среднее в каждой выборке. Гистограмма оценок этих средних показала бы их распределение (рис. 10-1); это — распределение выборочных средних. Мы можем увидеть следующее.

- Если объём выборки разумно большой, оценки среднего имеют нормальное распределение при любом распределении исходных данных в популяции (это следует из теоремы, известной как центральная предельная теорема).
- Если объём выборки небольшой, оценки среднего отвечают нормальному распределению при условии, что данные в популяции также отвечают нормальному распределению.
- Среднее этих оценок — несмещённая оценка истинного среднего в популяции (генерального среднего), т.е. среднее этих оценок эквивалентно истинному среднему в популяции.
- Вариабельность распределения выражается стандартным отклонением оценок, известным как стандартная ошибка среднего (часто обозначают как Standard Error Means, SEM). Если бы мы знали стандартное отклонение популяции σ, тогда стандартная ошибка среднего описывалась бы так:

$$\text{SEM} = \frac{\sigma}{\sqrt{n}}.$$

- В случае если есть, как обычно, только одна выборка, нашей лучшей оценкой среднего популяции будет выборочное среднее, а так как редко бывает известно стандартное отклонение в популяции (генеральный стандарт), то стандартную ошибку среднего оценивают следующим образом:

$$\text{SEM} = \frac{s}{\sqrt{n}},$$

где s — стандартное отклонение наблюдений в выборке.

Стандартная ошибка среднего отражает точность нашей оценки.

ИНТЕРПРЕТАЦИЯ СТАНДАРТНОЙ ОШИБКИ

- Большая стандартная ошибка указывает, что оценка неточна.
- Небольшая стандартная ошибка указывает, что оценка точна.
- Стандартная ошибка уменьшится, т.е. мы получим более точную оценку, если:
  - ✧ объём выборки увеличится (рис. 10-1);
  - ✧ данные имеют небольшое рассеяние.

СТАНДАРТНОЕ ОТКЛОНЕНИЕ (SD) ИЛИ СТАНДАРТНАЯ ОШИБКА СРЕДНЕГО (SEM)?

Хотя на первый взгляд кажется, что эти два параметра схожи, их используют в разных целях. Стандартное отклонение отражает вариабельность в значениях данных и должно быть указано, если нужно пояснить изменчивость в наборе данных. Наоборот, стандартная ошибка отображает точность выборочного среднего и должна быть указана, если интересует среднее значение набора данных.

ВЫБОРОЧНОЕ РАСПРЕДЕЛЕНИЕ ПРОПОРЦИЙ

Нас может интересовать пропорция индивидуумов в популяции, которые обладают некоторыми специфичными особенностями. Если взять выборку объёма n из популяции, то лучшая оценка p пропорции в популяции π вычисляется следующим образом:

$$p = r/n,$$

где r — количество индивидуумов в выборке с характерными особенностями. Если повторно извлечь выборки объёмом n из популяции и нанести оценки пропорции на гистограмму, то исходное выборочное распределение пропорций почти отвечало бы нормальному распределению со средним значением π. Стандартное отклонение распределения оценённых пропорций — стандартная ошибка пропорции. Когда берут только одну выборку, то её оценка производится так:

$$\mathrm{SE}(p) = \sqrt{\frac{p(1-p)}{n}}.$$

Эта величина представляет собой показатель точности оценки π; малая стандартная ошибка отражает высокую точность оценки.

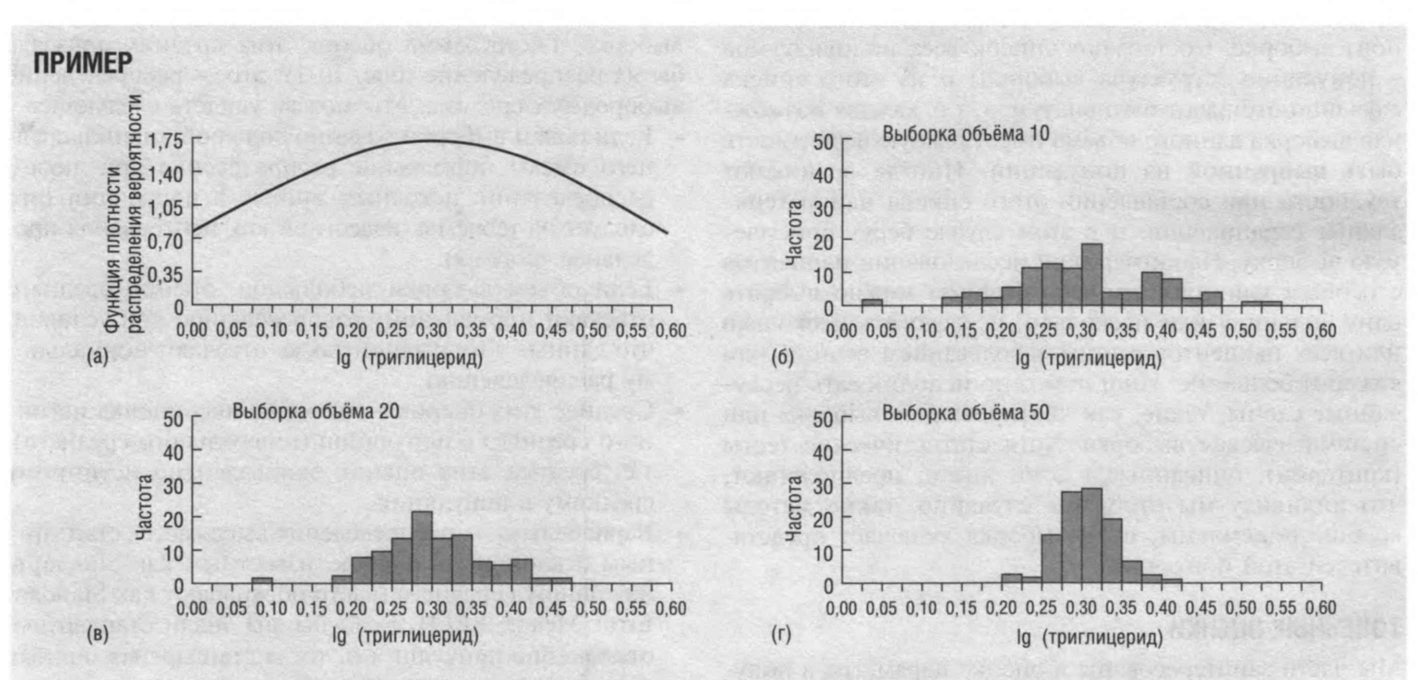


Рис. 10-1. (а) Теоретическое нормальное распределение величины lg триглицерида со средним =0,31 lg (ммоль/л) и стандартное отклонение =0,24 lg (ммоль/л) и наблюдаемые распределения средних 100 случайных выборок объёма (б) 10; (в) 20; (г) 50, взятые из этого теоретического распределения

11 Доверительные интервалы

Взяв выборку из популяции, мы получим точечную оценку интересующего нас параметра и вычислим стандартную ошибку для того, чтобы указать точность оценки. Однако для большинства случаев стандартная ошибка как таковая неприемлема. Гораздо полезнее объединить эту меру точности с интервальной оценкой для параметра популяции. Это можно сделать, используя знания о теоретическом распределении вероятности выборочной статистики (параметра) для того, чтобы вычислить доверительный интервал (ДИ) для параметра. Вообще, доверительный интервал расширяет оценки в обе стороны некоторой величиной, кратной стандартной ошибке (данного параметра); два значения (доверительные границы), определяющие интервал, обычно отделяют запятой и заключают в скобки.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ СРЕДНЕГО

Использование нормального распределения

Выборочное среднее $\bar{x}$ имеет нормальное распределение, если объём выборки большой (см. главу 10), поэтому можно применить знания о нормальном распределении при рассмотрении выборочного среднего. В частности, 95% распределения выборочных средних находится в пределах 1,96 стандартных отклонений (SD) среднего популяции. Когда у нас есть только одна выборка, мы называем это SEM и вычисляем 95% доверительного интервала для среднего следующим образом:

$$\bar{x} - (1{,}96 \times \text{SEM}); \quad \bar{x} + (1{,}96 \times \text{SEM}).$$

Если повторить этот эксперимент много раз, то интервал будет содержать истинное среднее популяции в 95% случаев. Обычно этот доверительный интервал объясняют как, например, интервал значений, в пределах которого с доверительной вероятностью 95% находится истинное среднее популяции (генеральное среднее). Хотя это не вполне строго (среднее в популяции есть фиксированное значение и поэтому не может иметь вероятность, отнесённую к нему) таким образом интерпретировать доверительный интервал (CI — Confidence Interval, ДИ — доверительный интервал), но концептуально это удобнее для понимания.

Использование *t*-распределения

Можно использовать нормальное распределение, если знать значение дисперсии в популяции. Кроме того, когда объём выборки небольшой, выборочное среднее отвечает нормальному распределению, если данные, лежащие в основе популяции, распределены нормально. Там, где данные, лежащие в основе популяции, распределены ненормально и/или неизвестна генеральная дисперсия (дисперсия в популяции), выборочное среднее подчиняется *t*-распределению Стьюдента. Вычисляем 95% доверительный интервал для генерального среднего в популяции следующим образом:

$$(\bar{x} - (t_{0,05} \times \text{SEM}); \quad \bar{x} + (t_{0,05} \times \text{SEM})), \text{ т.е. } \bar{x} \pm t_{0,05} \times \frac{s}{\sqrt{n}},$$

где $t_{0,05}$ — процентная точка (процентиль) *t*-распределения Стьюдента (табл. А2, см. приложение А) с $(n-1)$ степенями свободы, которая даёт двустороннюю вероятность 0,05. Вообще, она обеспечивает более широкий доверительный интервал, чем при использовании нормального распределения, поскольку учитывает дополнительную неопределённость, которую вводят, оценивая стандартное отклонение популяции и/или из-за небольшого объёма выборки. Когда объём выборки большой (порядка 100 и более), разница между двумя распределениями (*t*-Стьюдента и нормальным) незначительна. Тем не менее всегда используют *t*-распределение при вычислении доверительных интервалов, даже если объём выборки большой.

Обычно указывают 95% ДИ. Можно вычислить другие доверительные интервалы, например 99% ДИ для среднего. Вместо произведения стандартной ошибки и табличного значения *t*-распределения, которое соответствует двусторонней вероятности 0,05, умножают её (стандартную ошибку) на значение, которое соответствует двусторонней вероятности 0,01. Это более широкий доверительный интервал, чем в случае 95%, поскольку он отражает увеличенное доверие к тому, что интервал действительно включает среднее популяции.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ПРОПОРЦИИ

Выборочное распределение пропорций имеет биноминальное распределение. Однако если объём выборки n разумно большой, тогда выборочное распределение пропорции приблизительно нормально со средним π. Оцениваем π выборочным отношением $p=r/n$ (где r — количество индивидуумов в выборке с интересующими нас характерными особенностями), и стандартная ошибка оценивается:

$$\sqrt{\frac{p(1-p)}{n}},$$

95% доверительный интервал для пропорции оценивается:

$$\left(p - \left[1{,}96 \times \sqrt{\frac{p(1-p)}{n}}\right]; \quad p + \left[1{,}96 \times \sqrt{\frac{p(1-p)}{n}}\right]\right).$$

Если объём выборки небольшой (обычно когда $n \times p$ или $n \times (1-p)$ меньше 5), тогда необходимо использовать биноминальное распределение для того, чтобы вычислить точные доверительные интервалы. Заметьте, что если p выражается в процентах, то $(1-p)$ заменяют на $(100-p)$.

ИНТЕРПРЕТАЦИЯ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ

При интерпретации доверительного интервала нас интересуют следующие вопросы.

- Насколько широк ДИ? Широкий доверительный интервал указывает на то, что оценка неточна; узкий указывает на точную оценку. Ширина доверительного интервала зависит от размера стандартной ошибки, которая, в свою очередь, зависит от объёма выборки и при рассмотрении числовой переменной от изменчивости данных, поэтому исследования с небольшим набором данных дают более широкие доверительные интервалы, чем исследования многочисленного набора данных немногих переменных.
- Какой клинический смысл можно извлечь из ДИ? Верхние и нижние пределы дают оценку, будут ли результаты клинически значимы (см. пример).
- Включает ли ДИ какие-либо значения, представляющие особенный интерес? Можно проверить, ложится ли вероятное значение для параметра популяции в пределы доверительного интервала. Если да, то результаты согласуются с этим вероятным значением. Если нет, тогда маловероятно (для 95% доверительного интервала шанс почти 5%), что параметр имеет это значение.

СТЕПЕНИ СВОБОДЫ

В статистике вам обязательно встретится термин «степени свободы» (*df*). В общем, их можно вычислить как объём выборки минус число ограничивающих условий в отдельном конкретном вычислении; эти ограничения могут быть параметрами, которые необходимо оценить. В качестве простого примера рассмотрим набор из трёх чисел, которые суммируются в конкретную сумму T. Два из них «свободны», чтобы принять любое значение, а оставшееся число фиксировано единственным ограничением, введённым T, поэтому числа имеют две степени свободы. Точно так же вычисляются степени свободы выборочной вариации:

$$s^2 = \frac{\sum (x-\overline{x})^2}{n-1}.$$

(См. главу 6) объём выборки минус один, потому что мы должны вычислить выборочное среднее ($\overline{x}$), оценку среднего популяции, для того чтобы оценить s^2.

ПРИМЕР

Доверительный интервал для среднего

Мы заинтересованы в определении среднего возраста женщин при первых родах, у которых наблюдаются беспорядочные кровотечения. На выборке 49 таких женщин (см. главу 2).

- Средний возраст при рождении ребёнка, $\overline{x}$ = 27,01 лет.
- Стандартное отклонение, s = 5,1282 лет.
- Стандартная ошибка среднего, SEM=5,1282/$\sqrt{49}$ = 0,7326 лет.

Переменная приблизительно нормально распределена, поскольку дисперсия в популяции неизвестна, используем *t*-распределение для того, чтобы вычислить доверительный интервал. 95% доверительный интервал для среднего это:

$$27{,}01 \pm (2{,}011 \times 0{,}7326) = (25{,}54;\ 28{,}48) \text{ лет},$$

где 2,011 — процентная точка *t*-распределения с (49−1)=48 степенями свободы, которая даёт двустороннюю вероятность 0,05 (приложение А, табл. А2).

Мы на 95% уверены, что истинный средний возраст женщин при первых родах с беспорядочными кровотечениями в популяции колеблется от 25,54 до 28,48 лет. Этот интервал довольно узкий, отражает точную оценку. В общей популяции в 1997 г. средний возраст при первых родах был 26,8 лет. Так как 26,8 попадает в наш доверительный интервал, то это небольшое свидетельство того, что женщины с беспорядочными кровотечениями имеют тенденцию рожать в более старшем возрасте, чем другие женщины.

Заметьте, что 99% доверительный интервал (25,05; 28,97 лет) немного шире, чем 95% ДИ, отражает наше увеличенное доверие, что среднее в популяции расположено в этом интервале.

Доверительный интервал для пропорции

Из 64 исследуемых женщин 27 (42,2%) заявили, что у них кровоточили дёсны по крайней мере один раз в неделю. Это относительно высокий процент и может обеспечить метод для идентификации недиагностированных женщин с беспорядочными кровотечениями в основной популяции. Вычислим 95% доверительный интервал для пропорции с кровоточащими дёснами в популяции.

Выборочная пропорция равна: 27/64=0,422.

Стандартная ошибка пропорции равна:

$$\sqrt{\frac{0{,}422(1-0{,}422)}{64}} = 0{,}0617.$$

95% доверительный интервал равен:

$$0{,}422 \pm (1{,}96 \times 0{,}0617) = (0{,}301;\ 0{,}543).$$

Мы на 95% уверены, что истинный процент женщин с беспорядочными кровотечениями в популяции, у которых кровоточат дёсны, колеблется в частотном интервале от 30,1 до 54,3%. Это довольно широкий доверительный интервал, предлагающий недостаточную точность; значительно больший объём выборки дал бы возможность получить более точную оценку. Однако обе границы (верхняя и нижняя) доверительного интервала указывают на то, что существенный процент этих женщин по всей вероятности испытывают кровотечение дёсен. Прежде чем делать какие-либо выводы относительно значения этого факта для идентификации недиагностированных женщин с беспорядочными кровотечениями, необходимо было бы получить оценку частоты такого недомогания в основной популяции.

12 План исследования I

План исследования жизненно важен, так как плохо разработанные исследования могут ввести в заблуждение и дать неправильные результаты. Огромное количество данных, полученных при несбалансированном исследовании, не компенсирует ошибки, имеющиеся в проекте. В этой главе и главе 13 мы обсудим некоторые главные аспекты проекта для исследований. В главах 14–16 мы рассмотрим особые типы исследований: клиническое, когортное и исследование типа «случай–контроль».

Цели любого исследования должны быть чётко определены вначале. Мы можем захотеть оценить параметр в популяции (как, например, риск некоторых явлений) для того, чтобы рассмотреть связь между этиологическим фактором и интересующим нас исходом или чтобы оценить влияние вмешательства (такого, как новое лечение). Существует ряд возможных планов для любого такого исследования. Окончательный выбор плана будет зависеть не только от целей, но и от имеющихся в распоряжении денежных средств и этических соображений (табл. 12-1).

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИЛИ ИЗУЧЕНИЕ НАБЛЮДЕНИЕМ

- Экспериментальное исследование подразумевает участие исследователя, который будет влиять некоторым образом на результат. Клиническое исследование — пример экспериментального исследования, при котором исследователь привносит какую-либо форму лечения. Другие примеры включают исследования животных или лабораторные исследования, которые выполняют на основе эксперимента. Экспериментальное исследование предоставляет наиболее убедительные доказательства для любой гипотезы, поскольку появляется возможность контролировать факторы, которые могут повлиять на результат. Однако эти исследования не всегда выполнимы или, если вовлечены люди или животные, могут быть неэтичными.
- Исследование — результат наблюдения, например, когортное или исследование «случай–контроль». Исследование, в котором исследователь ничего не делает, чтобы повлиять на результат, а просто наблюдает за тем, что происходит; предоставляют ещё более скудный объём информации, чем экспериментальные исследования, потому что часто невозможно контролировать все факторы, которые воздействуют на результат. Однако в некоторых случаях они становятся единственно полезными и возможными способами исследования. Эпидемиологические исследования, которые оценивают соотношение между факторами и заболеваниями в популяции, относятся к наблюдательным.

Оценка причинности в наблюдательных исследованиях

Хотя наиболее убедительное доказательство для причинной роли фактора при заболевании обычно получают из экспериментальных исследований, информацию из наблюдательных исследований можно использовать при условии, что она отвечает ряду критериев. Наиболее известные критерии для оценки причинной обусловленности были предложены Хиллом[1].

- Причина должна предшествовать результату.
- Взаимосвязь должна быть правдоподобной, т.е. результаты должны быть биологически разумны.

РЕЗУЛЬТАТЫ РЯДА ИССЛЕДОВАНИЙ ДОЛЖНЫ БЫТЬ СОВМЕСТИМЫ

- Связь между причиной и результатом должна быть достаточно сильная.
- Должно быть соотношение между ответной реакцией на дозу и результатом, т.е. верхние уровни воздействия должны привести к более тяжёлому заболеванию или к более быстрому началу заболевания.
- Удаление фактора должно уменьшить риск заболевания.

ПЕРЕКРЁСТНОЕ (ПОПЕРЕЧНОЕ) ИЛИ ЛОНГИТЮДАЛЬНОЕ (ПРОДОЛЬНОЕ) ИССЛЕДОВАНИЕ

- Перекрёстные исследования выполняют в единственной временно́й точке. Примеры включают обзоры и переписи популяции. Особенно они подходят для точечной оценки распространённости состояния в популяции.

$$\text{Количество заболеваний} = \frac{\text{Точечная оценка распространённости в единственной временной точке}}{\text{Общее количеств, исследуемое в той же самой временной точке}}.$$

Поскольку неизвестно, когда именно до начала исследования произошли события, можно только сказать, что существует связь между интересующим нас фактором и заболеванием, а не то, что, по всей вероятности, именно этот фактор вызвал заболевание. Кроме того, мы не можем оценить частоту заболевания, т.е. частоту новых случаев в определённый период. К тому же так как перекрёстные исследования выполняют только в одной временной точке, то мы не можем рассматривать тренды спустя какое-то время. Однако эти исследования можно быстро и дёшево провести.

- Повторное перекрёстное исследование можно выполнять в различные периоды времени (временные точки) для того, чтобы оценить тренды через какое-то время. Однако, поскольку эти исследования охватывают различные группы пациентов в

[1] *Hill A.B.* The environment and disease: association or causation? Proceedings of the Royal Society of Medicine. — 1965. — Vol. 58. — P. 295.

3 Заказ 2051

каждой временно́й точке, бывает трудно оценить, отражают ли очевидные изменения через какое-то время различия в группах исследуемых пациентов.

- Лонгитюдальные исследования отвечают за выборку пациентов во времени. Обычно они проспективны в том плане, что пациентами занимаются ещё до какой-либо временной точки. Иногда ретроспективные исследования, при которых отбирают пациентов, и факторы из их прошлого идентифицируют так же, как лонгитюдальные. Лонгитюдальные исследования занимают больше времени, чем перекрёстные исследования, и, следовательно, требуют больших капиталовложений и, если они используют память пациента или медицинские записи, могут быть смещены (объяснение в конце этой главы).

Экспериментальные исследования обычно проспективные, так как рассматривают влияние вмешательства на будущие последствия. Однако изучение наблюдением может быть как проспективным, так и ретроспективным.

КОНТРОЛЬ

При планировании исследования и интерпретации каких-либо его результатов необходимо использовать сравнительную группу или контрольную группу. Например, при оценке причинной роли отдельного фактора на заболевание риск заболевания нужно рассматривать как на тех, кто подвергается, так и на тех, кто не подвергается рассматриваемому фактору. Смотрите также «Сравнения видов лечения» в главе 14.

СМЕЩЕНИЕ

Когда есть систематическое различие между результатами исследования и истинным положением дел, происходит, как считают, смещение.

Типы смещения.

- Смещение наблюдателя. Один наблюдатель недостаточно последовательно или чрезмерно подробно сообщает об отдельной переменной.
- Смещение, обусловленное сопутствующими факторами. Возникает подложная ассоциация вследствие отказа полностью установить факторы, которые связаны как с риском, так и с результатом.
- Смещение отбора. Больные, отобранные для исследования, не из той популяции, к которой будут применять результаты.
- Информационное смещение (обусловленное измерением). Измерения систематически неправильно записывались.
- Публикационное смещение. Тенденция публиковать только те статьи, в которых содержатся положительные и актуальные результаты.

Есть и другие смещения. Например, ошибка субъективности ответа, или ошибка памяти, эффект здорового участника, смещение оценки и распределения.

Таблица 12-1. Планы исследования

Тип исследования	Временной период	Форма	Действие в прошлом	Действие в настоящее время (отправная точка)	Действие в будущем	Типичное использование
Перекрёстный	Перекрёстный	Наблюдательная		Сбор всей информации		• Оценка распространённости. • Референтные интервалы и дагностические тесты • Текущее состояние здоровья в группах
Повторное перекрёстное	Перекрёстный	Наблюдательная		Сбор всей информации	Сбор всей информации	• Влияние изменения времени
Когортное (см. главу 15)	Проспективное	Наблюдательная		Определение. Когорты и оценка факторов риска → следствие	Наблюдение результата	• Прогноз и естественное развитие (с больным или болезнью) • Этиология
Случай–контроль (см. главу 16)	Ретроспективное	Наблюдательная	Оценка факторов риска	← проследить: Определение основной группы и группы контроля		• Этиология (особенно для редких болезней)
Экспериментальное	Проспективное	Экспериментальная		Применение воздействия → следствие	Наблюдение результата	• Клиническое испытание для оценки терапии • Испытание для оценки превентивного измерения, например вакцинация в больших масштабах • Лабораторный эксперимент

13 План исследования II

ДИСПЕРСИЯ

Дисперсия (вариация) в данных может быть вызвана известными факторами, ошибками измерений или может быть необъяснённой случайной вариацией. Мы измеряем влияние вариации в наборе исходных данных на оценку параметра популяции, используя стандартную ошибку. Когда измерение переменной подчинено значительной вариации, оценки, касающиеся этой переменной, будут неточными, с большими стандартными ошибками. Совершенно понятно, что желательно уменьшить влияние вариации насколько это возможно и таким образом увеличить точность оценок. Для этого есть различные способы.

Повторности

Оценки будут более точными, если сделать повторные измерения (например, два или три измерения данной переменной для каждого пациента по каждому наблюдению). Однако, как ни копируй, измерения не всегда независимы, поэтому мы должны быть внимательны при анализе таких данных. Простой подход состоит в том, чтобы использовать среднее каждого набора повторений при анализе вместо оригинальных измерений. В качестве альтернативы можно использовать методы, которые определённо учитывают повторения измерений.

Объём выборки

Выбор соответствующего объёма для исследования — решающий фактор в плане исследования. При увеличенном объёме выборки стандартная ошибка оценки будет сокращаться и приведёт к увеличенной точности и мощности исследования. Вычисления объёма выборки должны быть выполнены до начала исследования.

Особенные планы исследования

Модификация простых планов исследования может привести к более точным оценкам. По существу, мы сравниваем эффект одного или более «методов лечения» на экспериментальных единицах. Экспериментальная единица — самая маленькая группа «пациентов», которую можно принимать как независимую для целей анализа, например, один больной, объём крови или кусочек кожи. Если экспериментальные единицы назначены случайно для лечения и нет никаких других усовершенствований согласно плану, тогда мы имеем полностью рандомизированный план. Хотя этот план нужно непосредственно анализировать, это будет неэффективно, если есть существенная вариация между экспериментальными единицами. В данном случае можно объединить наблюдения в блок и/или использовать план исследования с перекрытием для того, чтобы уменьшить воздействие этой вариации.

Блокирование. Часто используют для экспериментальных единиц, которые выделяют по соответствующим признакам в однородный блок или слой (например, такие блоки могут представлять различные возрастные группы). Вариация между единицами в одном блоке меньше, чем между единицами в разных блоках. В каждом блоке пациентов выбирают для лечения случайно; мы предпочитаем сравнивать виды лечения в пределах каждого блока, чем делать полное сравнение между пациентами в различных блоках: можно оценить эффекты лечения точнее, чем если бы не было блокирования.

Параллельные планы исследования против перекрёстных планов (рис. 13-1). Вообще, мы сравниваем пациентов в различных группах. Например, большинство клинических испытаний — это параллельные испытания, когда больной проходит один или два (а иногда больше) курса лечения, которые сравниваются, т.е. они приводят к сравнениям между пациентами.

Так как при измерении одного пациента вариация обычно меньше, чем между разными пациентами, то в некоторых случаях предпочтительнее рассматривать каждого пациента как его собственный контроль. Эти сравнения данных одного пациента предоставляют более точные результаты, чем сравнения данных разных пациентов, и для исследования нужно меньшее количество пациентов, чтобы достичь того же самого уровня точности. При проведении клинических испытаний план исследования с перекрытием[1] — пример сравнения данных одного пациента. Если проводится два курса лечения, то каждый пациент проходит каждый курс лечения, один за другим в произвольном порядке для того, чтобы устранить любое влияние календарного периода. Периоды лечения разделяют периодом «размыва», который позволяет рассеять любые остаточные эффекты (эффекты наложения) предыдущего лечения. Анализируем разницу в реакциях на эти два лечения для каждого пациента. Этот план можно использовать только тогда, когда лечение временно облегчает симптомы, а не обеспечивает излечение, и период реакции не очень длительный.

ФАКТОРНЫЕ ЭКСПЕРИМЕНТЫ[2]

Когда мы заинтересованы более чем в одном факторе воздействия, то раздельные исследования, которые оценивают эффект изменения одного фактора одновременно, но порознь, могут быть неэффективными и дорогостоящими. Факторные планы позволяют совокупно анализировать любое количество рассматри-

---

[1] Senn S. Cross–over Trials in Clinical Research. Wiley, Chichester, 1993.

[2] В середине XX в. в прикладной статистике выделилось даже отдельное направление — математическая теория планирования экспериментов. Согласно основным постулатам этой теории допускается одновременное варьирование многих факторов по строго установленным правилам. Преимущества такого подхода заключаются в том, что многократно снижаются затраты на экспериментальные исследования и значительно повышается точность оценки параметров уравнений связи. Много ссылок на книги по этому направлению читатели могут найти по адресам http://www.biometrica.tomsk.ru/razdel_2_1.htm — http://www.biometrica.tomsk.ru/razdel_2_12.htm Прим. переводчика.

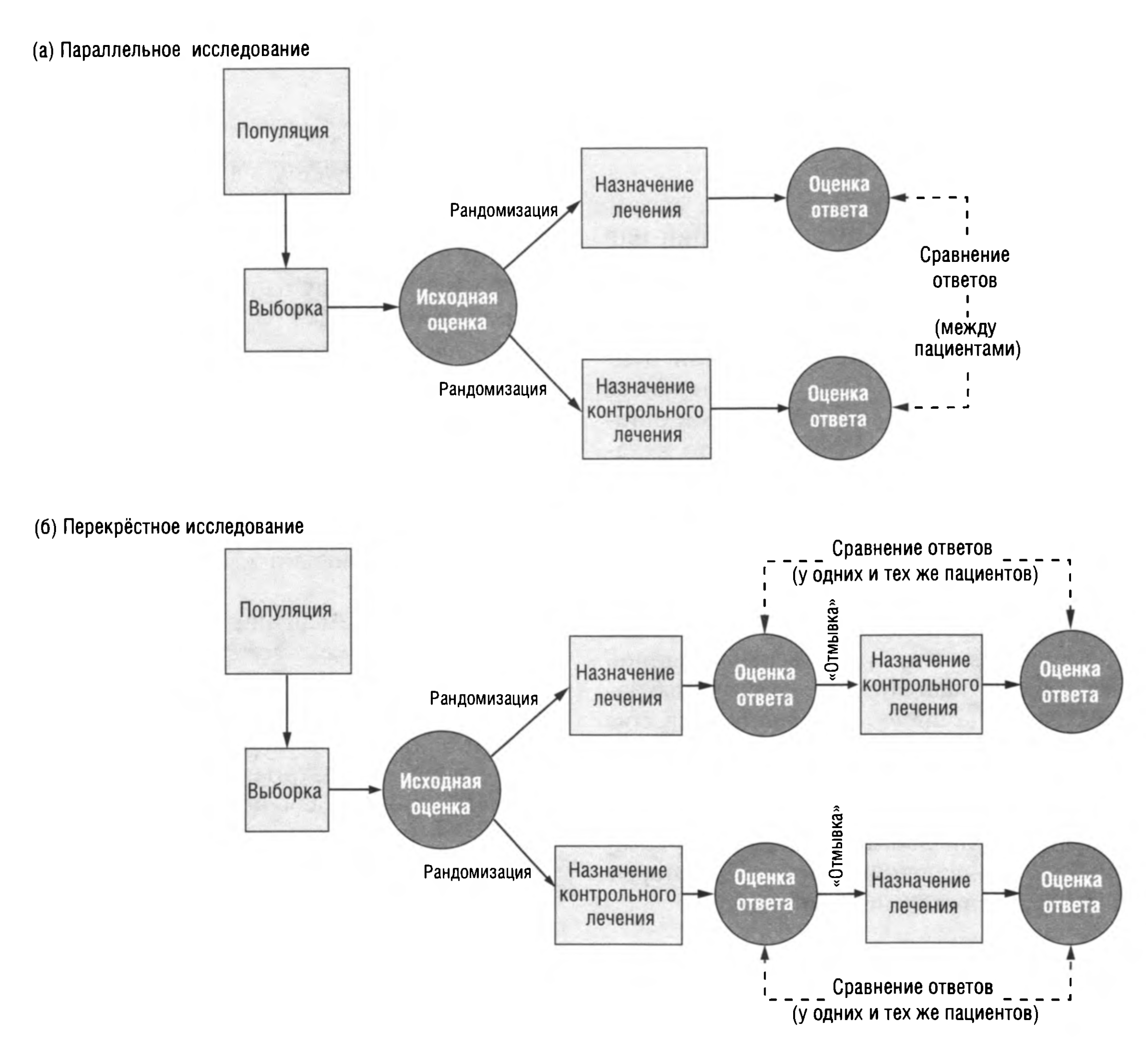


Рис. 13-1. Параллельный и перекрёстный планы исследования.

ваемых факторов. Самый простой план, факторный эксперимент 2×2, рассматривает два фактора (например, два различных лечения), каждый на двух уровнях (например, либо активное, либо пассивное лечение). Рассмотрим исследование американских врачей[1], разработанное для того, чтобы оценить всю важность аспирина и бета-каротина для предупреждения болезни сердца и рака. Использовали два лекарственных состава: активное лекарство и плацебо, и два уровня: применял врач или нет каждый из составов. Табл. 13-1 показывает возможные комбинации этих факторов.

Таблица 13-1. Возможные комбинации лечения

	β-Каротин	
Аспирин	Нет	Да
Нет	Ничего	β-каротин
Да	Аспирин	Аспирин и β-каротин

Оцениваем эффект уровня бета-каротина, сравнивая данные пациентов в левой и правой колонках. Точно так же оцениваем эффект уровня аспирина, сравнивая данные пациентов в верхней и нижней строках. Кроме того, можно проверить, есть ли взаимодействие между этими двумя факторами, т.е. когда эффект уровня бета-каротина отличается на каждом из двух уровней аспирина. И тогда мы можем сказать, что есть взаимодействие между этими двумя факторами. Взаимодействие наводило бы на мысль, что сочетание аспирина и бета-каротина было бы более (или менее) эффективно, чем ожидалось при простом суммировании отдельных эффектов каждого лекарственного средства. Этот план предоставляет дополнительную информацию для двух раздельных исследований и более эффективно использует денежные средства, требуя меньший объём выборки для того, чтобы получить оценки с данной степенью точности.

[1] Steering Committee of the Physician's Health Study Research Group. Final report of the aspirin component of the on–going Physicians Health Study // New England Journal of Medicine, 1989. — Vol. 321. — P. 129–135.

14 Клинические испытания

Клинические испытания — это любая форма запланированного экспериментального исследования, разработанного для того, чтобы оценивать клинические результаты нового лечения. Клинические исследования могут быть или доклиническими изучениями — незначительные клинические изучения для исследования эффекта и безопасности (I/II фаза исследований), или полными оценками нового лечения (III фаза исследований). В этой главе мы обсудим главные аспекты III фазы исследований, о которых следует сообщать в любом издании (см. CONSORT сообщение, табл. 14-1, рис. 14-1 и 14-2).

Таблица 14-1. Контрольный список разделов для включения в отчёт о рандомизированном контролируемом испытании

Глава документа и тема	№	Описание	Стр.
Заголовок и резюме	1	Каким образом участники распределялись на опытную и контрольную группы (например, «случайное распределение», «рандомизированное» или «случайный выбор лечения»)	
Введение, обоснование	2	Научные предпосылки и обоснование проблемы	
Методы, участники	3	Критерии включения участников и учреждений и территорий, где проводился сбор данных	
Вмешательства	4	Точные детали вмешательств, назначенных для каждой группы (как и когда они фактически осуществлялись)	
Цели	5	Конкретные цели и гипотезы	
Результаты	6	Ясно определить критерии (первостепенных и второстепенных исходов) и по возможности любые методы, применявшиеся для повышения качества измерений (например, многократные наблюдения, обучение исследователей)	
Объём выборки	7	Как был определён объём выборки и по возможности объяснение условий промежуточного анализа, а также условий прекращения исследования	
Рандомизация (генерация последовательности)	8	Метод, использованный для генерации случайных чисел, включая ограничительные детали рандомизации (например, блокирование, расслоение)	
Рандомизация (маскировка распределения)	9	Метод, обеспечивающий исполнение случайного характера распределения участников (например, пронумерованные контейнеры, телефонная станция), с указаниями, сохранялась ли информация о распределении участников по группам вмешательств в тайне до момента назначения лечения	
Выполнение рандомизации	10	Кто разрабатывал последовательность распределения, занимался регистрацией участников и кто назначал участников в группы?	
«Ослепление» (маскирование)	11	Действительно ли распределение участников и результаты вмешательства подвергались процедуре маскирования? Когда есть возможность описать каким образом была оценена успешность процедуры «ослепления»?	
Статистические методы	12	Статистические методы, использованные для сравнения групп по первичным исходам. Методы дополнительного анализа, включая анализ подгрупп и корректирующий анализ	
Результаты, движение участников	13	Прохождение участников каждой стадии исследований (настоятельно рекомендуется диаграмма, см. рис. 14-1). Конкретно указать численность участников, рандомизированно назначенных в каждую группу и получивших определённый вид лечения, завершивших исследование согласно протоколу, включённых в анализ первичных результатов. Опишите отклонения от протокола плановых исследований вместе с причинами отклонений	
Комплектование	14	Даты, определяющие период включения пациентов и проведение исследования	
Основные данные (о пациентах)	15	Основные демографические и клинические характеристики каждой группы	
Численности анализов	16	Численность пациентов в каждой группе (знаменатель), включённых в каждый анализ, проводился ли анализ в зависимости от назначенного лечения (первоначального распределения рандомизированных участников без включения). Выразите результаты в абсолютных числах, если возможно (например, 10/20, но не 50%)	
Исходы и оценка	17	Для каждого первичного и вторичного исхода привести итоговые результаты для каждой группы и оценки выраженности эффекта, указывая объём наблюдений и точность оценок (например, 95% доверительный интервал)	
Вспомогательные анализы	18	Разнообразные ссылки любых других выполненных исследований, включая исследования анализа подгрупп, скорректированные исследования, указывающие те же условия и исследования	
Неблагоприятные исходы	19	Все важные неблагоприятные исходы или побочные эффекты в каждой группе вмешательства	
Обсуждения и интерпретация	20	Интерпретация результатов, принимая во внимание гипотезы исследования, потенциальные источники смещения или неточности результатов, опасности, связанные со сложностью анализов и исходов	
Обобщаемость	21	Обобщаемость (внешняя валидность, достоверность) результатов испытаний	
Всеобъемлющие доказательства	22	Общая интерпретация результатов в контексте имеющихся доказательств	

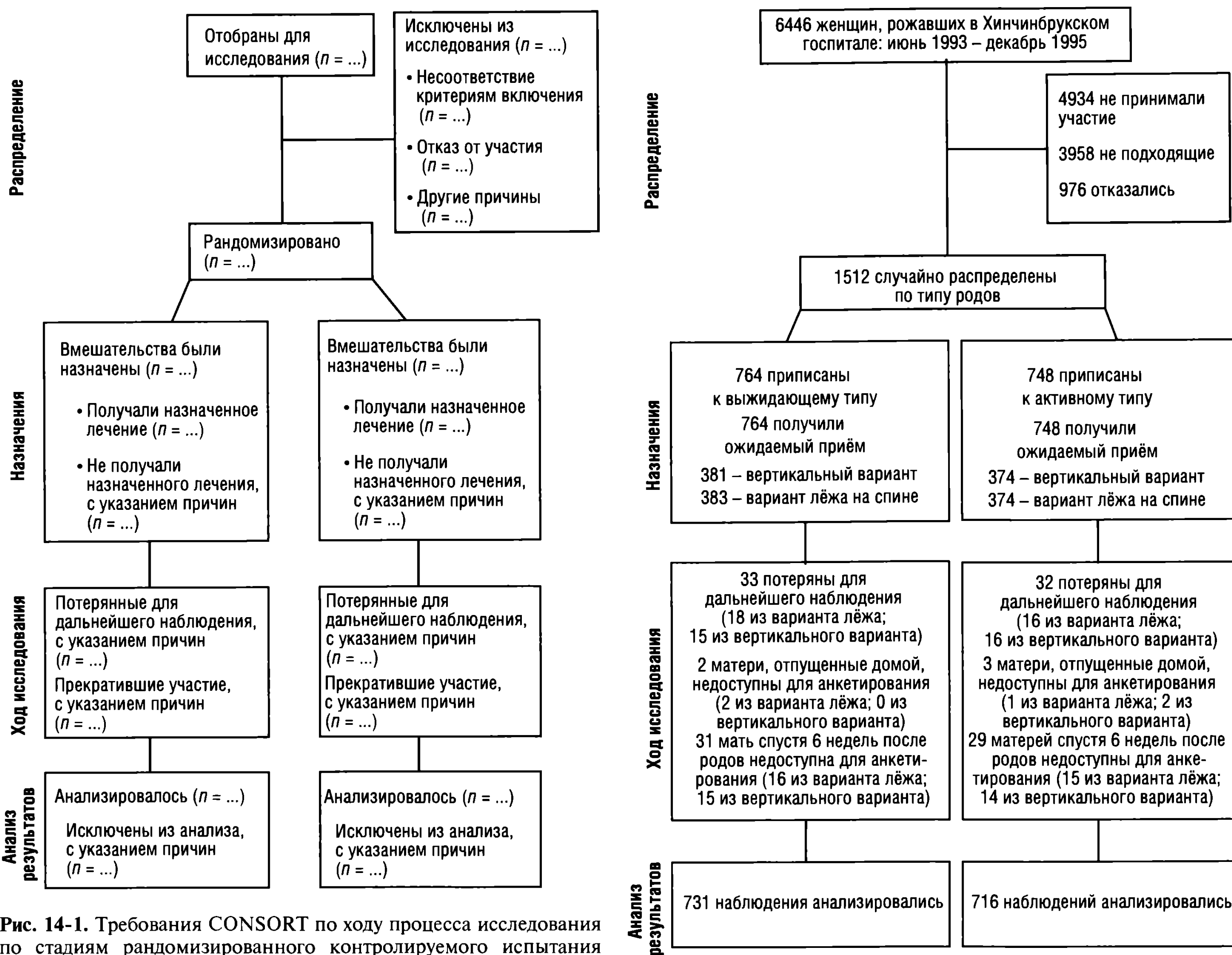


Рис. 14-1. Требования CONSORT по ходу процесса исследования по стадиям рандомизированного контролируемого испытания (http://www.consort–statement.org). См. русскоязычную версию по адресу: http://www.consort–statement.org/mod_product/uploads/CONSORT%20Statement%202001_Russian.pdf

Рис. 14-2. Примерный профиль испытания (адаптировано из описания испытания из главы 40 с разрешения авторов).

ЛЕЧЕНИЕ СО СРАВНЕНИЕМ

Клинические исследования — проспективные исследования, в которых мы заинтересованы при измерении воздействия данного лечения в настоящее время, чтобы использовать возможный результат в будущем. Вообще, клинические испытания оценивают новое вмешательство (например, тип или дозу лекарственного средства или хирургические процедуры). По всему разделу мы допускаем для простоты, что выбранное новое лечение необходимо оценивать.

Важная особенность клинического исследования — то, что оно должно быть сравнительным, сопоставимым. Без контрольного лечения невозможно быть уверенным в том, что любая реакция — единственно надлежащий эффект лечения и значение нового лечения может быть завышено. Контролем может быть стандартное лечение (положительный контроль) или, если такового не существует, плацебо (средство, которое и по виду, и по вкусовым качествам похоже на новое лекарственное средство, но не содержит ни одного активного компонента) — отрицательный контроль, или отсутствие лечения, если этические соображения это позволяют.

КОНЕЧНЫЕ ТОЧКИ

Нужно решить заранее, какой исход наиболее точно отражает преимущество новой терапии. Исход, известный как первичная конечная точка исследования, обычно касается эффективности лечения. Вторая конечная точка, которая часто касается токсичности, также представляет интерес и должна быть рассмотрена вначале. Анализ всех конечных точек проводят в конце исследования. Однако можно выполнить заранее запланированный промежуточный анализ (например, чтобы быть уверенными, что никакой токсичности, из-за которой пришлось бы остановить исследование, не наблюдается). Следует проявить осторожность при сравнении видов лечения в эти периоды из-за проблемы многократного испытания гипотезы.

РАСПРЕДЕЛЕНИЕ ЛЕЧЕНИЯ

Как только больного формально вводят в клинические испытания, его распределяют в группу лечения. Вообще, больных распределяют, используя процесс, известный как случайное распределение или рандомизация, часто при помощи созданного компьютерного

списка случайных чисел или используя таблицу случайных чисел (приложение А, табл. А12). Например, чтобы распределить больных в две группы лечения, можно использовать последовательность случайных чисел и определить больного в группу лечения А, если это число чётное, и в группу лечения В, если оно нечётное. Этот процесс порождает сходство между группами лечения в выражении основных характеристик при вступлении в испытание (т.е. он помогает избежать смещения распределения), следовательно, смешивание максимизирует эффективность испытания.

Если основная характеристика исследования не распределена однородно в группах лечения, изучаемые оценки соответствующих итоговых мер, например средних и стандартных отклонений, будут различаться более вероятно, чем при использовании рандомизированного распределения в группы. Будет непригодной формальная проверка статистической гипотезы (например, с помощью *t*-критерия), чтобы сравнить любые параметры основной характеристики в группах лечения, потому что испытание гипотезы оценивает, является ли различие между группами случайным.

Испытания, в которых больные рандомизируются для получения либо нового лечения, либо контрольного лечения, называют рандомизированными контролируемыми испытаниями (часто упоминают как РКИ) и расценивают как оптимальные.

Также существуют дальнейшие усовершенствования рандомизации, включая стратифицированную рандомизацию (которая контролирует эффекты важных факторов) и блоковую рандомизацию (которую обеспечивают приблизительно равные по объёму группы лечения). Систематического размещения, посредством которого больных систематически распределяют в группы лечения либо в зависимости от дня посещения, либо по дате рождения, следует избегать по мере возможности. Врач может назначить предложенное лечение для отдельного больного прежде, чем его введут в испытание, и это может повлиять на его решение относительно того, включать ли больного в испытание. Иногда используют процесс, известный как групповая рандомизация, посредством которого насколько возможно произвольно распределяют группы пациентов (например, все люди, зарегистрированные согласно общепринятой практике) для лечения, а не отдельного пациента. В таких проектах следует обратить особое внимание на планирование объёма исследования и анализ данных.

«ОСЛЕПЛЕНИЕ», ИЛИ МАСКИРОВАНИЕ

Когда больной и/или врачи осведомлены о распределении в группы лечения, может произойти смещение оценки. Знание распределения вида лечения может повлиять на регистрацию признаков улучшения или неблагоприятных явлений, поэтому, если возможно, все участвующие (врачи, больные, эксперты) в испытании должны быть «ослеплены» относительно распределения пациентов по группам лечения. Испытание, в котором ни больной, ни врач/эксперт не осведомлены о распределении лечения, называют дважды слепым испытанием. Испытания, при которых невозможно «ослепить» больного, называют испытаниями с однократным ослеплением (маскированием) при условии, что врач и/или оценщик «ослеплены» относительно распределения лечения.

ЭТИЧЕСКИЕ ПРОБЛЕМЫ

Так как клинические испытания связаны с людьми, то проблемы больного имеют значение. Любое клиническое испытание должно быть одобрено этическим комитетом, который принимает решение относительно того, не нарушает ли данное испытание Хельсинкскую декларацию. Согласие должно быть получено от всех больных, прежде чем их введут в испытание.

ПРОТОКОЛ

До начала любого клинического испытания должно быть подготовлено письменное описание всех аспектов испытания — протокол. Он включает в себя информацию относительно целей и задач испытания наряду с определением того, какие больные будут участвовать (критерий включения и исключения), график лечения, сбор данных и анализ, план на случай непредвиденных обстоятельств и список персонала, принимающего участие в исследовании. Очень важно привлечь достаточное количество больных, в этом случае появится возможность точнее определить истинный эффект лечения, поэтому, прежде чем начать испытание, необходимо высчитать оптимальный объём испытания.

Отклонения от протокола — больные, которые участвуют в испытании, но не выполняют критерии протокола. Например, больные, которые неправильно привлечены к участию в исследовании или отозваны, больные, которые переведены на другое лечение. Чтобы избежать смещения, не мешало бы проанализировать исследование на основе стремления лечиться: вся информация анализируется в тех группах, куда больные первоначально были распределены, независимо от того, придерживались ли они режима лечения. По возможности следует попытаться собрать информацию о больных, которых исключили из испытания. Анализ пролеченных, при котором учитывают только больных, закончивших полный курс лечения, не рекомендуется, поскольку он часто приводит к смещённым сравнениям лечения.

15 Когортные исследования

При когортном исследовании берётся группа пациентов и обычно отслеживается во времени. Цель исследования — определить, воздействует ли отдельный этиологический фактор на будущий результат заболевания (рис. 15-1). Если да, то это фактор риска для исхода заболевания. Например, в ряде когортных исследований изучали зависимость между диетическими факторами и раком. Хотя большинство когортных исследований — проспективные, можно исследовать исторические когорты, полученная при этом информация будет ретроспективна. Однако качество исторических исследований часто зависит от медицинских отчётов (записей) и поэтому они могут подвергаться смещению.

Когортные исследования могут быть либо фиксированными, либо динамическими. Если пациенты покидают фиксированную когорту, то их не заменяют. В динамических когортах одни пациенты могут исчезнуть из когорты, другие — присоединиться, поскольку они становятся подходящими.

ВЫБОР КОГОРТЫ

Когорта должна быть представителем популяции, для которой будут обобщаться получаемые результаты. Часто она имеет преимущества, если пациентов будут набирать из такого источника, как отдельная профессиональная группа (например, государственные служащие, медицинские практикующие врачи), поскольку информацию относительно их смертности и заболеваемости можно просто получить из документации, имеющейся на рабочем месте, и с пациентами можно опять контактировать при необходимости. Однако такая когорта не может быть истинным представителем основной популяции, а может быть более здоровая. Также когорты могут быть набраны из списков общего назначения при условии, что принимать участие в исследовании будет группа пациентов с различным состоянием здоровья. Однако есть тенденция того, что эти больные будут иметь схожие социальные корни, потому что они проживают в одном и том же районе.

Чтобы оценить этиологический результат фактора риска, пациенты, набранные в когорту, должны быть свободными от болезни в начале исследования. Это будет гарантией того, что любая подверженность фактору риска произойдет до получения результата, таким образом допуская, что причинная роль фактора будет постулированной. По причине того что в начале исследования пациенты свободны от болезни, мы часто наблюдаем эффект здорового участника. Процент смертности в первый период исследования может быть гораздо ниже, чем был бы в общей популяции. Это станет очевидным, когда процент смертности начнёт со временем увеличиваться (период исследований может длиться несколько лет).

ПОСЛЕДУЮЩЕЕ НАБЛЮДЕНИЕ ЗА ПАЦИЕНТАМИ

При наблюдении за пациентами в течение длительного времени всегда есть опасность, что кто-то не будет участвовать в продолжении исследования. Пациенты могут переехать, не оставив нового адреса, или захотят выйти из исследования. Если большое количество пациентов будет потеряно для продолжения исследования, то когортные исследования не принесут пользы. Таким образом, мы должны найти способ, чтобы свести к минимуму эти исчезновения, сохраняя постоянный контакт с пациентами.

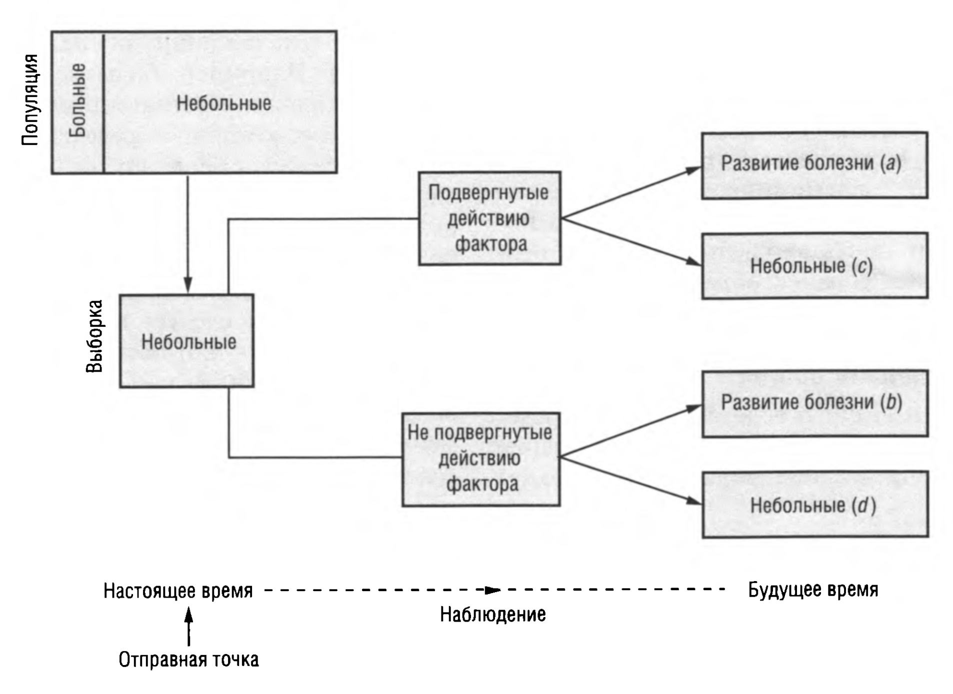


Рис. 15-1. Представление в виде диаграммы когортного исследования (повторение вводных слов см. табл. 15-1).

ИНФОРМАЦИЯ ОБ ИСХОДАХ И ВНЕШНИХ ВОЗДЕЙСТВИЯХ

Важно получить полную и точную информацию относительно результатов заболевания, например, о смертности и причинах болезни. Это повлечёт за собой просмотр записей болезни, статистики смертности, многоцелевых и больничных отчётов.

Подверженность рискам, представляющим интерес, может измениться на определённом этапе исследования. Например, оценивая соотношение между потреблением алкоголя и заболеваниями сердца, следует учитывать, что типичное потребление алкоголя пациентом, по всей вероятности, изменится через какое-то время, поэтому важно заново опрашивать пациентов, чтобы получить новые данные в изменившихся обстоятельствах.

АНАЛИЗ КОГОРТНЫХ ИССЛЕДОВАНИЙ

Табл. 15-1 содержит частоты наблюдений.

Таблица 15-1. Частота наблюдений (см. рис. 15-1)

	Подверженность фактору		
	Да	**Нет**	**Итого**
Заболевание, представляющее интерес			
Да	*a*	*b*	*a*+*b*
Нет	*c*	*d*	*c*+*d*
Итого	*a*+*c*	*b*+*d*	*n*=*a*+*b*+*c*+*d*

Так как за больными наблюдают лонгитюдально через какой-то период времени, то возможно оценить риск развития болезни в популяции путём вычисления риска в исследованной выборке.

Оценочный риск заболевания =

$$= \frac{\text{Число развития болезни через какой-то период исследования}}{\text{Итоговое число в когорте}} = \frac{a+b}{n}.$$

Риск заболевания у пациентов, подверженных и не подверженных фактору, представляющему интерес, в популяции можно оценить таким же способом.

Оценочный риск заболевания в подверженной группе:

$$\text{риск} = a/(a+c).$$

Оценочный риск заболевания в неподверженной группе:

$$\text{риск} = b/(b+d).$$

Затем оценочный относительный

$$\text{риск} = \text{риск}_{\text{подв}} \,/\, \text{риск}_{\text{неподв}} = a/(a+b){:}b/(b+d)$$

$$\frac{\text{риск}_{\text{гр. с воздействием}}}{\text{риск}_{\text{гр. без воздействия}}} = \frac{a/(a+c)}{b/(b+d)}.$$

Относительный риск (ОР) измеряет увеличенный (или уменьшенный) риск заболевания, связанный с подверженностью фактору, представляющему интерес. Относительный риск, равный 1, указывает, что существует тот же самый риск как в подверженной, так и неподверженной группах. Относительный риск больше чем 1 указывает, что существует увеличенный риск в подверженной группе по сравнению с неподверженной группой; относительный риск меньше чем 1 указывает на сокращение риска заболевания в подверженной группе. Например, относительный риск, равный 2, указывал бы, что пациенты в подверженной группе имели бы вдвое больше риск заболевания, чем пациенты в неподверженной группе.

Если вычислить и доверительные интервалы для относительного риска, тогда можно проверить, равен ли единице относительный риск. Это легко выполнить на компьютере, и поэтому мы опускаем подробности.

ПРЕИМУЩЕСТВА КОГОРТНЫХ ИССЛЕДОВАНИЙ

- Можно оценить время последовательности событий.
- Они обеспечивают информацией относительно размаха результатов.
- Можно непосредственно измерить частоту/риск заболевания.
- Можно собрать подробную информацию относительно подверженности широкому размаху факторов.
- Можно исследовать подверженность факторам, которые редко встречаются.
- Можно измерить подверженность в ряде временных точек так, чтобы можно было бы исследовать изменения в подверженности через какое-то время.
- Существуют уменьшенное смещение ошибки памяти и смещение выбора по сравнению с исследованиями «случай–контроль».

НЕДОСТАТКИ КОГОРТНЫХ ИССЛЕДОВАНИЙ

- Вообще, при когортных исследованиях за пациентами наблюдают в течение длительного периода времени, и поэтому они являются дорогостоящими.
- Там, где результат, представляющий интерес, редок, необходим очень большой объём выборки.
- По мере того как приближается окончание исследования, часто наблюдается увеличение потери больных, поскольку они могут переехать или выйти из исследования, что ведет к смещённым результатам.
- Как следствие длительного масштаба времени, часто бывает трудно сохранить последовательность измерений и результатов через какое-то время. Кроме того, пациенты могут изменить своё поведение после первой беседы.
- Возможно, что результаты заболевания и их вероятности или этиология заболевания как таковая могут измениться через какой-то период времени.

КЛИНИЧЕСКИЕ КОГОРТЫ

Иногда мы выбираем когорту пациентов с тем же самым клиническим состоянием, посещающих одну или более больниц, и следуем за ними как пациентами или амбулаторными больными, чтобы видеть, сколько пациентов имеют положительные исходы при наличии признаков усиления заболевания, например смерти или рецидива. Информация, которую мы собираем о каждом пациенте, собрана как часть обычного клинического наблюдения. Цели клинических когорт или

наблюдательных баз данных могут включать результаты лечения индивидуумов с описанием состояния и оценкой эффектов различных подходов к лечению (например, различные лекарства или различные способы лечебного воздействия). В отличие от рандомизированного распределения в контролируемых испытаниях, которые часто включают специально выбранных индивидуумов, тех, кто желает участвовать в таких испытаниях, клинические когорты часто включают всех пациентов с изучаемым в больнице болезненным состоянием. Таким образом, результаты от этих когорт, как думается, более точно отражают итоги, которые были бы замечены в клинической практике. Однако поскольку распределение по виду лечения в этих исследованиях неслучайно, то клинические когорты имеют склонность к смещению, вызванному сопутствующими факторами.

ПРИМЕР

Британское региональное обследование сердца — большое когортное исследование на 7735 мужчинах в возрасте от 40 до 59 лет, произвольно отобранных в 24 городах Британии, чтобы определить факторы риска ишемической болезни сердца. В процессе набора для обследования мужчин опросили относительно их образа жизни и демографических факторов, включая информацию о привычке курить. Из 7718 мужчин, которые предоставили информацию о курении, 5899 (76,4%) курили в определённый период жизни, включая тех, которые курят сейчас, и бывших курильщиков. В течение 10 лет 650 из 7718 мужчин (8,4%) перенесли инфаркт миокарда (ИМ). Результаты, отображённые в таблице, показывают количество (и процент) курящих и некурящих, у которых случился и не случился ИМ в течение 10 лет.

Таблица 15-2. Инфаркт миокарда за 10 лет

Статус курения	Да	Нет	Итого
Когда-либо курил	563 (9,5%)	5336 (90,5%)	5899
Никогда не курил	87 (4,8%)	1732 (95,2%)	1819
Итого	650 (8,4%)	7068 (71,6%)	7718

$$\text{Оценочный относительный риск} = \frac{(563/5899)}{(87/1819)} = 2{,}00.$$

Можно показать, что 95% доверительный интервал для истинного относительного риска составляет (1,6; 2,49).

Мы можем интерпретировать относительный риск, подразумевая таким образом, что у мужчины среднего возраста, который когда-либо курил, вдвое больше вероятность перенести инфаркт миокарда в течение следующих 10 лет, чем у мужчины, который никогда не курил. В качестве альтернативы риск перенести ИМ для мужчины, который когда-либо курил, на 100% больше, чем для мужчины, который никогда не курил.

Данные любезно предоставлены Dr. F.C. Lampe, Ms. M. Walker and Dr. p. Whincup. Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK.

16 Исследование «случай–контроль»

Исследование «случай—контроль» сравнивает характерные особенности группы больных с особым исходом болезни (случай) с группой пациентов без исхода болезни (контроль) для того, чтобы определить насколько чаще (или менее часто) встречаются какие-либо факторы в случае, чем в контроле (рис. 16-1). Такие ретроспективные исследования не предоставляют информацию относительно распространённости или частоты заболевания, но могут предоставить о заключения относительно того, какие факторы увеличивают или сокращают риск заболеваемости.

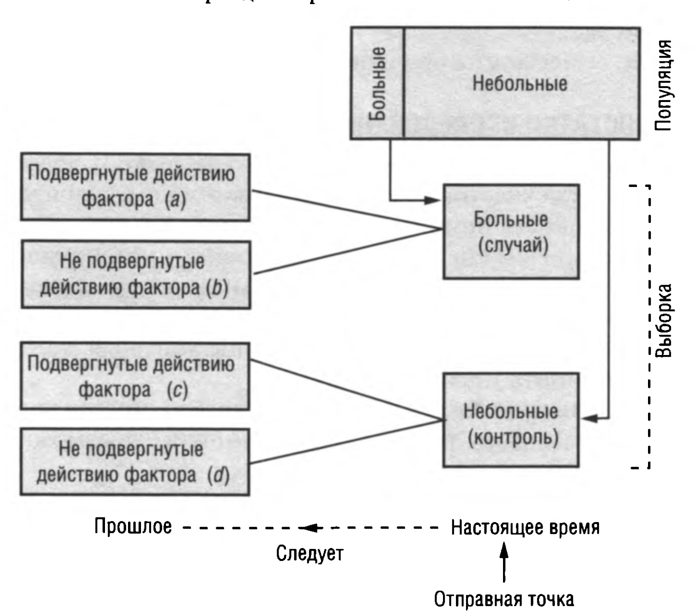


Рис. 16-1. Представление в виде диаграммы исследования «случай—контроль».

ВЫБОР СЛУЧАЕВ

Важно определить, нужно ли набирать число вновь обнаруженных больных (больные, которые были набраны во время диагноза) или ранее диагностированных пациентов (пациенты, которые уже были диагностированы до начала исследования). У ранее диагностированных пациентов, возможно, было время, чтобы отразить в их анамнезе подверженности факторам риска, особенно если заболевание хорошо известно, такое, как рак, или они могли изменить своё поведение после диагноза. Важно провести как можно больше наблюдений, чтобы результаты были более вескими, а заключения можно было бы применить к будущим популяциям. К концу, возможно, появится необходимость обратиться к спискам больницы и журналам регистрации болезней и включить тех, кто умер в течение того периода времени, когда были определены случай и контроль, потому что их исключение может привести к смещённой выборке заболеваний.

ВЫБОР КОНТРОЛЯ

Группу контроля до начала исследования необходимо предварительно проверить, чтобы быть уверенными, что у них нет болезни, представляющей интерес в данном исследовании. Иногда может быть более одного контроля для каждого случая. Если возможно, контроль следует выбирать из того же самого источника, что и случай. Очень часто контроль выбирают в больнице. Однако поскольку факторы риска, относящиеся к одному исходу болезни, могут относиться к другим исходам болезни, то выбор контроля, основанного на больнице, может проводиться из пациентов, которые уже подверглись фактору риска, представляющему интерес, и поэтому может быть не всегда соответствующим. Часто приемлемо осуществлять выбор из основной популяции, хотя такие пациенты не могут быть настолько заинтересованы в принятии участия в исследовании, и поэтому уровень реакции может быть ниже в контроле, чем в случае. Использование контроля из окружения больных может гарантировать, что случаи и контроли будут происходить из одного и того же социального слоя.

ПОДБОР

Большинство исследований «случай—контроль» согласовывают для того, чтобы выбрать случаи и контроли настолько схожими, насколько только это возможно. Вообще, это применяется при подборе пациентов по полу (т.е. если случай мужского пола, то и контроль должен быть мужского пола), но иногда пациентов можно подбирать по возрасту. Однако важно не подбирать на основе фактора риска, представляющего интерес, или на основе какого-либо другого фактора, который встречается в пределах причинной траектории болезни, поскольку будет невозможным оценить соотношение между фактором риска и заболеванием при исследовании. К сожалению, подбор всё-таки подразумевает, что невозможно исследовать эффект по данному заболеванию, используя переменные, которые применялись для подбора.

АНАЛИЗ НЕСОГЛАСОВАННЫХ ИССЛЕДОВАНИЙ «СЛУЧАЙ-КОНТРОЛЬ»

Таблица 16-1 показывает наблюдаемые частоты. Так как больных отбирают по состоянию заболевания, то невозможно оценить абсолютный риск заболевания. Мы можем вычислить отношение шансов следующим образом:

$$\text{Отношение шансов} = \frac{\text{Шансы существования случая в неподвергнутой группе}}{\text{Шансы существования случая в подвергнутой группе}}.$$

Таблица 16-1. Частота наблюдений (см. рис. 16-1)

	Подвергшиеся фактору		
	Да	**Нет**	**Итого**
Состояние болезни			
Случай	*a*	*b*	$a+b$
Контроль	*c*	*d*	$c+d$
Итого	$a+c$	$b+d$	$n=a+b+c+d$

Например, шансы существования случая в группе с фактором риска равны:

- вероятности существования случая в группе с фактором риска;
- вероятности отсутствия случая в группе с фактором риска.

Шансы существования случая в группе с фактором риска и в группе без фактора риска:

$$\text{шансы заболевания}_{\text{с фактором риска}} = \frac{\left(\dfrac{a}{a+c}\right)}{\left(\dfrac{c}{a+c}\right)} = \frac{a}{c},$$

$$\text{шансы заболевания}_{\text{без фактора риска}} = \frac{\left(\dfrac{b}{b+d}\right)}{\left(\dfrac{d}{b+d}\right)} = \frac{b}{d}.$$

И, следовательно,

$$\text{оценочное отношение шансов} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}.$$

Когда частота заболевания редка, то отношение шансов — это оценка относительного риска, и интерпретируется подобным образом, т.е. она указывает на увеличенный (или уменьшенный) риск, связанный с подверженностью фактору, представляющему интерес. Отношение шансов, равное единице, указывает, что существует один и тот же риск в подверженной и неподверженной группах. Отношение шансов больше 1 указывает, что существует увеличенный риск в группе с фактором риска по сравнению с группой без фактора риска и т.д. Доверительные интервалы и критерии проверки гипотез также могут быть сгенерированы и для отношения шансов.

АНАЛИЗ СОГЛАСОВАННЫХ ИССЛЕДОВАНИЙ «СЛУЧАЙ-КОНТРОЛЬ»

Если возможно, в анализе согласованных исследований «случай–контроль» необходимо учитывать тот факт, что случай и контроль связаны друг с другом в результате подбора. Далее все детали методов анализа для согласованных исследований вы можете найти в Breslow и Day.

ПРЕИМУЩЕСТВА ИССЛЕДОВАНИЯ «СЛУЧАЙ-КОНТРОЛЬ»

- Они относительно быстры, дёшевы и просты в выполнении.
- Они особенно подходят для редких заболеваний.
- Можно исследовать огромный спектр факторов риска.
- Не существует потерь выбывания из исследования.

НЕДОСТАТКИ ИССЛЕДОВАНИЯ «СЛУЧАЙ-КОНТРОЛЬ»

- Смещение, возникающее вследствие ошибок памяти, когда участники исследования имеют различную способность помнить некоторые подробности относительно своего анамнеза, становится потенциальной проблемой. Например, больной раком лёгкого может хорошо помнить случайный период, когда он курил, тогда как больной из группы контроля может не помнить подобный период.
- Если начало болезни предшествовало подверженности фактору риска, то причинность невозможно вывести.
- Исследования «случай–контроль» не пригодны, если подверженность фактору риска редка.

ПРИМЕР

1327 женщин в возрасте от 50 до 81 года с переломами бедра, которые жили в густонаселённом городском районе в Швеции, участвовали в несогласованном исследовании «случай–контроль». Их сравнили с 3262 пациентами (контрольная группа) в пределах того же возраста, произвольно отобранными из государственного регистра. Мы сконцентрировали наше внимание на установлении факта, насколько существенно терапия замены гормона в постменопаузе (HRT) сокращает риск перелома бедра. Результаты, указанные в таблице, показывают количество женщин, которые использовали терапию замены гормона в постменопаузе, и тех, которые никогда не использовали или использовали ранее терапию замены гормона в постменопаузе в группах случая и контроля.

Наблюдаемое отношение шансов =
= (40×3023)/(239×1287)=0,39.

Можно показать, что 95% доверительный интервал для отношения шансов (0,28; 0,56).

Таблица 16-2. Исследование терапии замены гормона в постменопаузе (HRT) сокращает риск перелома бедра

	Текущий пользователь HRT	Никогда не принимали HRT/ ранее принимали HRT	Итого
С переломом бедра (случаи)	40	1287	1327
Без перелома бедра (контроли)	239	3023	3262
Итого	279	4310	4589

Женщина в постменопаузе в этом возрасте в Швеции, которая принимает HRT, таким образом, имела 39% риска перелома бедра по сравнению с женщиной, которая никогда не принимала или прежде принимала HRT, т.е., будучи настоящим пользователем HRT, сократила риск перелома бедра на 61%.

Данные заимствованы из: K. Michaelsson, J.A. Baron, B.Y. Farahmand et al. Hormone replacement therapy and risk of hip fracture: population based case–control study // British Medical Journal, 1998 — Vol. 316 — P. 1858–1863.

17 Проверка гипотез

Часто собирают выборочный набор данных для того, чтобы оценить, сколько существует аргументов против определённой гипотезы, сформулированной относительно популяции (генеральной совокупности). Используют процесс, известный как проверка гипотез (или проверка значимости), чтобы определить количественную меру уверенности против определённой гипотезы.

В этой главе описывается характер проверки гипотезы в целом (врез 17-1); подробности проверки конкретных гипотез рассматриваются в следующих главах. В качестве удобной ссылки каждая проверка гипотез содержится в блоке, оформленном аналогичным образом.

Врез 17-1. Проверка гипотез. Общий обзор

Мы установили пять стадий проверки гипотез.

- 1. Определить нулевую и альтернативную гипотезу при исследовании.
- 2. Отобрать необходимые данные из выборки пациентов.
- 3. Вычислить значение статистики критерия, отвечающей H_0.
- 4. Сравнить значение статистики критерия со значениями из известного распределения вероятности (для данной статистики).
- 5. Интерпретировать достигнутый уровень значимости *P*-значения и результаты.

ОПРЕДЕЛЕНИЕ НУЛЕВОЙ И АЛЬТЕРНАТИВНОЙ ГИПОТЕЗЫ

Мы всегда проверяем нулевую гипотезу (H_0), которая отвергает эффект (к примеру, разница в средних равна нулю) в популяции. Например, если мы заинтересованы в сравнении показателей курения у мужчин и женщин в популяции, нулевая гипотеза была бы:

H_0 — показатели курения одинаковые у женщин и мужчин в популяции.

Затем мы определяем альтернативную гипотезу (H_1), которая принимается, если нулевая гипотеза неверна. Альтернативная гипотеза в большей степени относится к той теории, которую мы хотели бы исследовать. Итак, на этом примере мы могли бы иметь:

H_1 — показатели курения различны у женщин и мужчин в популяции.

Мы не уточнили разницу в показателях курения, т.е. не установили, имеют ли мужчины более высокие или более низкие показатели, чем женщины в популяции. Такой подход известен как двусторонний критерий, потому что учитывается любая возможность, и он рекомендуется, поскольку мы редко бываем уверенными заранее в направлении какого-либо различия, если таковое существует. В некоторых, более редких, случаях можно использовать односторонний критерий для гипотезы H_1, когда направление эффекта задано или проводится проверка эквивалентности или снижения эффекта (см. последний раздел в этой главе).

ПОЛУЧЕНИЕ СТАТИСТИКИ КРИТЕРИЯ

Когда данные собраны, подставляем значения из выборки в формулу, используемую для критерия, который применяют, чтобы определить значение статистики критерия. Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы — обычно чем больше значение статистики критерия, игнорируя его знак, тем сильнее эти аргументы.

ПОЛУЧЕНИЕ *P*-ЗНАЧЕНИЯ

Все статистики критерия подчиняются известным теоретическим распределениям вероятности. Мы связываем значение статистики критерия, полученное из выборки, с уже известным распределением, которому она подчиняется, чтобы получить *P*-значение, площадь обоих (или иногда одного) «хвостов» распределения вероятности. Большинство компьютерных пакетов обеспечивают автоматическое вычисление двустороннего *P*-значения. *P*-значение есть вероятность получения нашего вычисленного значения критерия или ещё большего значения, если нулевая гипотеза истинна. Нулевая гипотеза всегда относится к популяции, представляющей для нас интерес больший, нежели выборка. Следовательно, нулевая гипотеза может быть либо истинной, либо ложной, и мы не можем интерпретировать *P*-значение как вероятность того, что нулевая гипотеза истинна.

ПРИМЕНЕНИЕ *P*-ЗНАЧЕНИЯ

Мы должны принять решение относительно того, сколько аргументов потребуется для того, чтобы была возможность решить, отвергнуть ли нулевую гипотезу в пользу альтернативной. Чем меньше *P*-значение, тем сильнее аргументы против нулевой гипотезы.

Традиционно мы полагаем, что если *P*-значение меньше 0,05, то имеется достаточно аргументов для того, чтобы отвергнуть нулевую гипотезу, хотя есть небольшой шанс появления результатов, если нулевая гипотеза истинна. Тогда мы отвергнем нулевую гипотезу и скажем, что результаты значимы на 5% уровне (рис. 17-1).

Напротив, если *P*-значение больше 0,05, то обычно мы приходим к выводу, что аргументов недостаточно для того, чтобы отвергнуть нулевую гипотезу. Мы не отвергаем нулевую гипотезу и можем заявить, что результаты не значимы на 5% уровне (рис. 17-1). Это совсем не означает, что нулевая гипотеза истинна; просто у нас нет достаточных аргументов, чтобы отвергнуть её.

Выбор 5% произволен. Если брать за основу 5%, то мы поступим неправильно, отвергнув нулевую гипотезу, когда она истинна. Неправильное отклонение нулевой гипотезы может привести к серьёзным последствиям, в этом случае необходимо потребовать более веских аргументов, прежде чем отвергнуть нулевую гипотезу, например, мы можем выбрать *P*-значение



Рис. 17-1. Функция плотности распределения вероятности для статистики критерия проверки двусторонней гипотезы, P=0,05.

0,01 (или 0,001). Выбранную границу отсечки (например, 0,05 или 0,01) называют **критическим уровнем** значимости критерия.

Определение результата только как значимого на определённом уровне граничного значения (например P<0,05) может ввести в заблуждение. Допустим, если P=0,04, то мы отвергнем H_0; однако если P=0,06, то мы не отвергнем её. Действительно ли они различны? Мы рекомендуем всегда указывать точное P-значение, обычно получаемое из результата компьютерного анализа[1].

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

Критерий проверки гипотез, который основан на знании закона распределения вероятности величины критерия и вычислен по набору данных, известен как **параметрический критерий**. Часто данные не подчиняются предположениям, которые лежат в основе этих методов. В этих случаях можно использовать непараметрические критерии (иногда упоминаемые как критерии, свободные от распределения, или ранговые методы). Эти критерии обычно заменяют данные их рангами (т.е. числа 1, 2, 3 и т.д., описывающие их положение в упорядоченном наборе данных) и не делают никаких предположений относительно распределения вероятности, которому отвечают данные.

Непараметрический критерий особенно пригоден, когда объём выборки небольшой (настолько, что невозможно было бы оценить закон распределения данных) и когда данные выражаются в категориальной шкале. Однако непараметрические критерии вообще неэкономно используют имеющуюся информацию; следовательно, они обладают меньшей мощностью (см. главу 18) в обнаружении реального эффекта, чем эквивалентный параметрический критерий, если будут удовлетворены все предположения, лежащие в основе параметрического критерия. Кроме того, они являются, прежде всего, критериями значимости, которые часто не предоставляют оценок интересующего нас эффекта; они, скорее, приводят к решениям, чем к оценке и пониманию данных.

КАКОЙ КРИТЕРИЙ ИСПОЛЬЗОВАТЬ?

Решение, какой статистический критерий использовать, зависит от плана исследования, типа переменной и распределения, которому отвечают исследуемые данные. Блок-схема, приведённая на обложке книги, поможет принять решение.

ПРОВЕРКА ГИПОТЕЗ ПРОТИВ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ

Доверительные интервалы и проверка гипотез близко связаны. Первоначальная цель проверки гипотезы состоит в том, чтобы принять решение и предоставить точное P-значение. Доверительные интервалы определяют количественно интересующий эффект (например, разницу в средних) и дают возможность оценить клинические значения результатов. Однако так как они предоставляют интервал вероятных значений для истинного эффекта, то их также можно использовать для принятия решения, если даже не предоставлены точные P-значения. Например, если бы гипотетическое значение для данного эффекта (например, значение, равное нулю) находилось вне 95% доверительного интервала, тогда мы полагали бы, что гипотетическое значение неправдоподобно, и отвергли бы H_0. В этом случае мы бы узнали, что уровень значимости P меньше 0,05, но не узнали бы его точное значение.

ТЕСТИРОВАНИЕ ЭКВИВАЛЕНТНОСТИ И НЕХУДШЕГО КАЧЕСТВА

В большинстве рандомизированных контролируемых испытаний с двумя или более различными стратегиями лечения нас обычно интересует демонстрация превосходства по крайней мере одного метода лечения над другим (другими). Однако в некоторых ситуациях можно полагать, что новый метод лечения (например, препарат) не будет клинически более эффективен, чем существующее лечение, но будет иметь другие преимущества, возможно, уменьшение побочных эффектов, количества таблеток или затрат. Тогда можно просто показать, что эффективность нового лечения подобна (в тестировании эквивалентности) или не существенно худшая (в тестировании нехудшего качества), чем существующего лечения.

При выполнении тестирования эквивалентности или тестирования нехудшего качества гипотеза, показывающая, что эффекты этих двух методик лечения являются одинаковыми, не соответствует действительности. Это потому, что (1) незначащий результат не подразумевает нехудшего качества/эквивалентность, и (2) даже если статистически существенный эффект обнаружен, это может быть клинически не столь важно. Вместо этого мы по существу полностью изменяем нулевую и альтернативную гипотезы в проверке эквивалентности так, чтобы нулевая гипотеза выразила различие, а альтернативная гипотеза выразила эквивалентность.

[1] Очень важная и актуальная для отечественных исследователей рекомендация! В большинстве публикаций авторы указывают либо p <0,05, либо p >0,05. Такие записи могут свидетельствовать либо о крайне низком уровне качества проведённого статистического анализа, либо вообще об отсутствии такового. Прим. переводчика.

Вернее, чем вычисление тестовой статистики, мы вообще приближаемся к проблеме оценки эквивалентности и не ухудшения качества, определяя, находится ли доверительный интервал для интересующего нас эффекта (например, различие в средних между двумя группами лечения) полностью или частично в пределах предопределённого диапазона эквивалентности (т.е. диапазон значений, определённых клиническими экспертами, который соответствует отсутствию эффекта клинической важности). Если весь доверительный интервал для интересующего эффекта находится в пределах диапазона эквивалентности, то мы заключаем, что эти два лечения эквивалентны; в этой ситуации, даже если верхние и самые низкие границы доверительного интервала говорят, что есть преимущества одного лечения над другим, это вряд ли будет иметь какое-то клиническое значение. В тестировании нехудшего качества мы хотим показать, что новое лечение не существенно хуже, чем стандартное (если бы новое лечение оказалось лучше, чем стандартное, то это была бы дополнительная награда!). В этой ситуации, если самая нижняя граница соответствующего доверительного интервала не находится ниже самой низкой границы диапазона эквивалентности, мы заключаем, что новое лечение не худшего качества.

Если иначе не определено, проверка альтернативной гипотезы в последующих главах — испытания превосходства. Заметьте, что методы для определения объёма выборки, описанные в главе 36, не обращаются к эквивалентности или проверке нехудшего качества. Объём выборки, требуемый для тестирования эквивалентности или нехудшего качества, вообще больший, чем для аналогичного тестирования превосходства, если все факторы, которые затрагивают объём выборки (например, уровень значимости, мощность), те же самые.

18 Ошибки при проверке гипотез

ПРИНЯТИЕ РЕШЕНИЯ[1]

Большинство проверяемых гипотез в медицинской статистике сравнивают между собой группы людей, которые подвергаются различным исследованиям. Мы можем быть заинтересованы, например, в сравнении эффективности двух форм лечения, чтобы сократить пятилетнюю смертность от рака груди. Для данного исхода (например, смерть) мы называем данное сравнение, представляющее интерес, (например, различные показатели смертности через 5 лет) интересующим нас эффектом, или, если уместно, эффектом лечения. Мы выражаем нулевую гипотезу как отсутствие эффекта (например, пятилетняя смертность от рака груди та же самая в двух группах, получающих разное лечение); двусторонняя альтернативная гипотеза — различие эффектов не равно нулю. Мы осуществляем критериальную проверку гипотезы, которая даёт нам возможность принять решение относительно того, достаточно ли аргументов, чтобы отвергнуть нулевую гипотезу. Можно принять только одно из двух решений: либо отвергнуть нулевую гипотезу и тогда принять альтернативную, либо нет.

ПРИНЯТИЕ НЕПРАВИЛЬНОГО РЕШЕНИЯ

Несмотря на то что мы надеемся сделать правильный вывод относительно нулевой гипотезы, нужно признать тот факт, что мы можем принять и неправильное решение, когда отвергаем/не отвергаем нулевую гипотезу, потому что у нас есть только выборочная информация. Возможные ошибки показаны в табл. 18-1.

Таблица 18-1. Возможные ошибки при проверке гипотез

	Отклонить H_0	Принять H_0
H_0 верна	Ошибка 1-го рода	Нет ошибки
H_0 ошибочна	Нет ошибки	Ошибка 2-го рода

Ошибка 1-го типа (рода)

Мы отвергаем нулевую гипотезу, когда она истинна, и делаем вывод, что имеется эффект, когда в действительности его нет. Максимальная вероятность допустить ошибку 1-го типа обозначается α (альфа). Это уровень значимости критерия; мы отвергаем нулевую гипотезу, если наше P-значение меньше уровня значимости, т.е. если $P < \alpha$.

Мы должны принять решение относительно значения α прежде, чем соберём данные; обычно назначают условное значение 0,05, несмотря на то что можно выбрать более ограничивающее значение, такое, как 0,01. Шанс допустить ошибку 1-го типа никогда не превысит выбранный нами уровень значимости, скажем $\alpha=0,05$, так как мы отвергнем нулевую гипотезу, только если $P < 0,05$. Если мы обнаружим, что $P > 0,05$, то не отвергнем нулевую гипотезу и, следовательно, не допустим ошибки 1-го типа.

[1] Детальный анализ ошибок статистического анализа, допускаемых отечественными исследователями в медицине и биологии, читатели могут найти в статье [92]. Прим. переводчика.

Ошибка 2-го типа (рода)

Мы не отвергнем нулевую гипотезу, когда она ложна, и сделаем вывод, что нет эффекта, когда как в действительности он существует. Шанс допущения ошибки 2-го типа обозначается β (бета); а величина $(1-\beta)$ называется мощностью критерия. Следовательно, мощность — это вероятность отклонения нулевой гипотезы, когда она ложна; т.е. это шанс (обычно выраженный в процентах) обнаружить статистически значимый, реальный эффект лечения в выборке данного объёма.

В идеале мы хотели бы, чтобы мощность нашего критерия была 100%; однако приходится признать, что это невозможно, так как всегда есть опасность, хотя и незначительная, допустить ошибку 2-го типа. К счастью, известно, какие факторы влияют на мощность, и, таким образом, мы можем контролировать мощность критерия, рассматривая их.

МОЩНОСТЬ И СВЯЗАННЫЕ ФАКТОРЫ

Планируя исследование, необходимо знать мощность предложенного критерия. Имеет смысл начинать исследование, когда есть «хороший» шанс обнаружить клинически уместный эффект (под «хорошим» мы подразумеваем, что мощность должна быть по крайней мере 70–80%). И безответственно начинать клиническое испытание, у которого, скажем, только 40% вероятности обнаружить реальный эффект лечения, к тому же это бесполезная трата времени и денежных средств.

Ряд факторов имеют прямое отношение к мощности данного критерия.

- Объём выборки: мощность критерия увеличивается по мере увеличения объёма выборки. Это означает, что у большей выборки больше, чем у незначительной, возможностей обнаружить клинически важный эффект, если таковой существует. Когда объём выборки небольшой, у критерия может быть неадекватная мощность, чтобы обнаружить отдельный эффект. В главе 36 мы объясним, как выбрать объём выборки, принимая во внимание мощность. Эти методы также можно использовать для оценки мощности критерия для точно установленного объёма выборки.
- Вариабельность наблюдений: мощность увеличивается по мере того, как вариабельность наблюдений уменьшается (рис. 18-1).
- Интересующий эффект: мощность критерия больше для более высоких эффектов. Таким образом, критерий проверки гипотез имеет больше шансов обнаружить значительный реальный эффект, чем незначительный.
- Уровень значимости: мощность будет больше, если уровень значимости больше (это эквивалентно вероятности допущения ошибки 1-го типа (α), в то время как

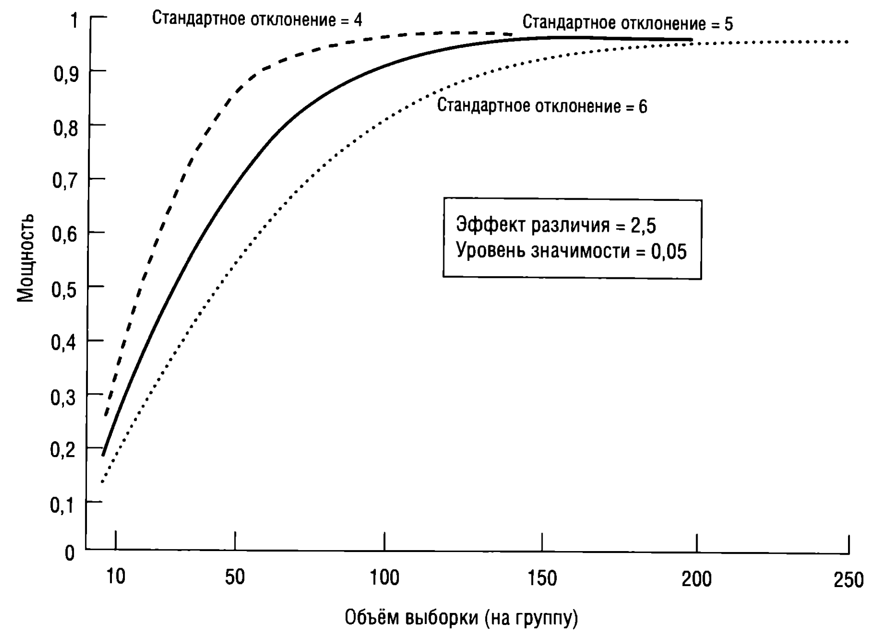


Рис. 18-1. Кривые мощности показывают соотношение мощности и объёма выборки в каждой из двух групп для сравнения двух средних, используя непарный *t*-критерий. Каждая кривая мощности связана с двусторонним критерием, для которого уровень значимости равен 0,05, а интересующий нас эффект различия средних (например, из-за различия в методах лечения) — 2,5. Предполагаемое одинаковое стандартное отклонение величин в двух группах различно для каждой кривой мощности (см. пример в главе 36).

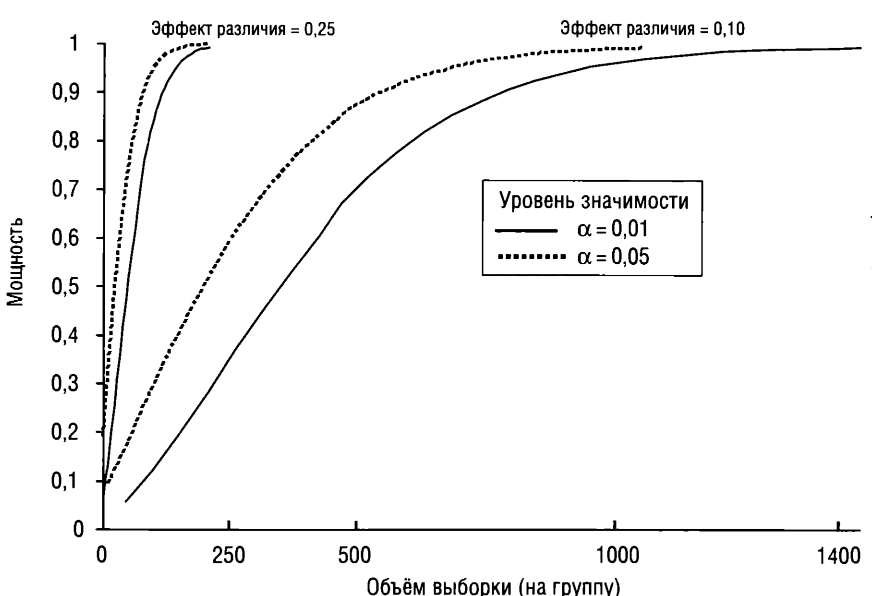


Рис. 18-2. Кривые мощности показывают соотношения между мощностью и объёмом выборки в каждой из двух групп для сравнения двух пропорций (долей), использующих критерий Хи-квадрат Пирсона. Кривые разведены друг от друга, когда интересующий нас эффект (т.е. разность пропорций интересующего нас свойства в двух группах лечения) равен или 0,25 (т.е. 0,65–0,4), или 0,1 (т.е. 0,5–0,4); уровень значимости двустороннего критерия равен или 0,05, или 0,01 (см. пример в главе 36).

ошибка 2-го типа (β) уменьшается). Таким образом, вероятнее всего мы обнаружим реальный эффект, если на стадии планирования решим, что будем рассматривать *P*-значение как значимое, если оно скорее будет меньше 0,05, чем меньше 0,01. На рис. 18-2 можно увидеть соотношение мощности и значимости.

Обратите внимание, что проверка доверительного интервала (см. главу 11) для интересующего нас эффекта указывает на то, была ли мощность адекватной. Большой доверительный интервал следует из небольшой выборки и/или набора данных с существенной вариабельностью и указывает на недостаточную мощность.

ПРОВЕРКА МНОЖЕСТВЕННЫХ ГИПОТЕЗ

Часто мы хотим выполнить критериальную проверку значимости множественных гипотез на наборе данных, например, когда он содержит много переменных или существует более двух видов лечения. Ошибка 1-го рода драматически увеличивается по мере увеличения количества сравнений, что приводит к ложным выводам относительно гипотез. Следовательно, нам нужно проверить только небольшое количество гипотез, выбранных для достижения первоначальной цели исследования и точно установленных априори. Можно использовать какую-нибудь форму апостериорного уточнения *P*-значения, принимая во внимание количество выполненных проверок гипотез. Например, при подходе Бонферрони (часто расцениваемый как довольно консервативный) умножаем каждое *P*-значение на количество выполненных проверок; тогда любые решения относительно значимости будут основываться на этом уточнённом *P*-значении.

4 Заказ 2051

19 Числовые данные: одна группа

ПРОБЛЕМА

Мы имеем выборку из одной группы пациентов и одну интересующую нас числовую или ординальную переменную. Мы интересуемся, принимает ли показатель среднего значения этой переменной некоторое определённое значение. Например, у нас может быть выборка пациентов со специфическим медицинским состоянием. Мы мониторировали уровень триглицеридов в крови здоровых пациентов и знаем, что они имеют геометрическое среднее 1,74 ммоль/л. Мы хотим знать, такой же ли средний уровень у наших пациентов с таким состоянием.

ОДНОВЫБОРОЧНЫЙ *t*-КРИТЕРИЙ

Допущения

В популяции переменная нормально распределена с данной (обычно неизвестной) дисперсией. Кроме того, мы имеем достаточный объём выборки, чтобы проверить допущение нормальности.

Обоснование

Мы интересуемся, отличается ли среднее значение μ в интересующей нас популяции от некоторой гипотетической величины μ_1. Применяем статистику критерия, которая основана на разнице между средней выборочной $\bar{x}$ и μ_1. Если дисперсия в популяции не известна, тогда эта статистика критерия, часто называемая t, подчиняется t-распределению Стьюдента. Если дисперсия в популяции известна или объём выборки очень большой, то можно использовать альтернативный критерий (часто называемый z-критерий), основанный на нормальном распределении. Однако в этих ситуациях результаты от обоих критериев по существу идентичны.

Дополнительное примечание

Выборка имеет размер n, а оценка стандартного отклонения равна s.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: средняя в популяции μ равна μ_1.
  - H_1: средняя в популяции не равна μ_1.
- 2. Отобрать необходимые данные из выборки пациентов.
- 3. Рассчитать величину статистики критерия, отвечающей H_0:

$$t = \frac{(\bar{x} - \mu_1)}{s/\sqrt{n}},$$

которая подчиняется t-распределению Стьюдента с $(n-1)$ степенями свободы.

- 4. Сравнить величину расчётной статистики t-критерия с величинами из табличного распределения вероятности t-критерия Стьюдента.
  - Обратиться к t-распределению в приложении А, табл. А2.
- 5. Интерпретировать величину P и результаты.
  - Интерпретируйте величину P и рассчитайте доверительный интервал для истинной (генеральной) средней популяции.
  - 95% доверительный интервал задается выражением:

$$\bar{x} \pm t_{0,05} \times \frac{s}{\sqrt{n}},$$

где $t_{0,05}$ — процентная точка t-распределения Стьюдента с $(n-1)$ степенями свободы для двусторонней вероятности в 0,05.

Интерпретация доверительного интервала

95% доверительный интервал даёт интервал значений, в котором, как мы на 95% уверены, лежит истинная средняя в популяции. Если 95% доверительный интервал не включает гипотетическую величину для средней μ_1, мы отбрасываем нулевую гипотезу при 5% уровне. Если, однако, доверительный интервал включает μ_1, то отбросить нулевую гипотезу при этом уровне значимости нельзя.

Если допущения не удовлетворены

Мы можем сомневаться, что переменная подчиняется нормальному распределению в популяции. В то время как t-критерий относительно устойчив к небольшому отклонению от нормальности, большее беспокойство должна вызывать асимметрия. Можно или преобразовать данные так, чтобы переменная была нормально распределена, или применить непараметрический критерий, такой, как знаковый критерий или критерий знаковых рангов Вилкоксона.

ЗНАКОВЫЙ КРИТЕРИЙ

Обоснование

Знаковый критерий — простой критерий, основанный на медиане распределения. Мы имеем некоторое гипотетическое значение λ для медианы в популяции. Если выборка извлечена из этой популяции, то около половины величин в ней должны быть более λ, а половина — менее λ (после исключения любых величин, равных λ). Знаковый критерий рассматривает ряд величин в выборке, которые больше (или меньше) чем λ.

Знаковый критерий — простой критерий; мы можем использовать более мощный критерий, например, критерий знаковых рангов Вилкоксона, который принимает во внимание как сами ранги данных, так и их знаки.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: медиана в популяции равна λ.
  - ✧ H_1: медиана в популяции не равна λ.
- 2. Отобрать необходимые данные из выборки пациентов.
- 3. Рассчитать величину статистики критерия, отвечающей H_0. Игнорируйте все значения, которые равны λ, в результате останется n' значений. Сосчитайте число значений, которые больше λ. Подобным же образом сосчитайте число значений, которые меньше λ. (На практике это часто включает расчёт разности между каждой величиной в выборке и λ, и учёт знака этой разности.) Положим, r есть меньшее из этих двух чисел.
  - ✧ Если $n' \le 10$, то статистика критерия есть r.
  - ✧ Если $n' > 10$, вычислим:

$$z = \frac{\left| r - \frac{n'}{2} \right| - \frac{1}{2}}{\frac{\sqrt{n'}}{2}},$$

где $n'/2$ — число значений выше (или ниже) медианы, которую мы должны ожидать, если бы нулевая гипотеза была верной.

Вертикальные линии означают, что мы берём абсолютное (т.е. положительное) значение чисел внутри линий. Распределение z приблизительно нормально. Вычитание $\frac{1}{2}$ в формуле для z — это поправка на непрерывность, которую мы должны включить, чтобы учесть, что мы относим дискретное значение (r) к непрерывному распределению (нормальному распределению).

- 4. Сравнить значение статистики критерия с величинами из известного распределения вероятности.
  - ✧ Если $n' \le 10$, описание r в приложении А, табл. А6.
  - ✧ Если $n' > 10$, описание z в приложении А, табл. А1.
- 5. Интерпретировать величину P и результаты. Интерпретируйте величину P и рассчитайте доверительный интервал для медианы — некоторые статистические программы обеспечивают это автоматически; если нет, мы можем ранжировать величины в пределах объёма выборки и обратиться к приложению А, табл. А7 для идентификации рангов значений, которые следует использовать для определения границ доверительного интервала. В целом доверительные интервалы для медианы будут больше, чем для среднего.

ПРИМЕР

Имеются данные, что высокие уровни триглицеридов в крови связаны с заболеванием сердца. Как часть большого когортного исследования сердечных заболеваний уровни триглицеридов были известны у 232 мужчин, у которых развилось сердечное заболевание в течение 5 лет после включения в исследование. Мы интересуемся, такие же ли средние уровни триглицеридов в популяции мужчин, из которых состоит эта выборка, как и в целой популяции. Для изучения этого был использован одновыборочный t-критерий. Уровни триглицеридов были асимметричны вправо (см. рис. 8-3, а); значения log (уровень триглицеридов) приблизительно нормально распределены (см. рис. 8-3, б), так что мы провели анализ на логарифмах уровней триглицеридов. У мужчин в целой популяции средний логарифм уровня триглицеридов равен 0,24 (lg ммоль/л), что эквивалентно геометрической средней в 1,74 ммоль/л.

- 1. H_0: средний log (уровень триглицеридов) в популяции мужчин, у которых развилось сердечное заболевание, равен 0,24 log (ммоль/л).
- 2. H_1: средний log (уровень триглицеридов) в популяции мужчин, у которых развилось сердечное заболевание, не равен 0,24 log (ммоль/л).
- 3. Объём выборки n=232; средний логарифм уровня триглицеридов: $\bar{x}$ равен 0,31 log (ммоль/л); стандартное отклонение логарифма уровня триглицеридов s=0,23 log (ммоль/л).
- 4. Статистика критерия:

$$t = \frac{0{,}31 - 0{,}24}{0{,}23/\sqrt{232}} = 4{,}64.$$

  - ✧ Обратившись к t-распределению в приложении А, табл. А2 с 231 степенями свободы, получим: $P<0{,}001$.
- 5. Имеются убедительные данные для отказа от нулевой гипотезы, что геометрическая средняя уровней триглицеридов в популяции мужчин, у которых развилось сердечное заболевание, оценивается как $10^{0,31}$ (2,04 ммоль/л). 95% доверительный интервал для геометрической средней уровней триглицеридов варьирует от 1,9 до 2,19 ммоль/л (т.е. антилогарифм $[0{,}31 \pm 1{,}96 \times 0{,}23/\sqrt{232}]$), поэтому в этой популяции пациентов геометрическая средняя уровня триглицеридов значительно выше, чем в общей популяции.

Можно использовать знаковый критерий для проведения подобного анализа на непреобразованных уровнях триглицеридов, поскольку он не задаёт каких-либо допущений о распределении. Мы допускаем, что медиана и среднее геометрическое уровней триглицеридов в мужской популяции подобны.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: медиана концентрации триглицеридов в популяции мужчин, у которых развилось сердечное заболевание, равна 1,74 ммоль/л.
  - H_1: концентрации триглицеридов в популяции мужчин, у которых развилось сердечное заболевание, не равна 1,74 ммоль/л.
- 2. В этой группе данных значение выборочной медианы равно 1,94 ммоль/л.
- 3. Мы исследуем разницу между каждым значением и 1,74. Есть 231 ненулевая разность, из которых 135 положительны и 96 отрицательны, поэтому $r = 96$. Поскольку число ненулевых разностей более 10, мы рассчитываем:

$$z = \frac{\left|96 - \frac{231}{2}\right| - \frac{1}{2}}{\frac{\sqrt{231}}{2}} = 2{,}50.$$

- 4. Обратившись к распределению z в приложении А, табл. А1, получим: $P = 0{,}012$.
- 5. Имеются данные за то, чтобы отбросить нулевую гипотезу, что медиана концентрации триглицеридов в популяции этих мужчин равна 1,74 ммоль/л. Формула в приложении А, табл. А7 показывает, что 95% доверительный интервал для медианы популяции задан посредством 101-й и 132-й ранжированных величин; это 1,77 и 2,16 ммоль/л, поэтому в данной популяции пациентов медиана концентрации триглицеридов статистически значимо выше, чем таковая в общей популяции.

Данные любезно предоставлены Ms F.C. Lampe, Ms M. Walker and Dr p. Whincup, Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK.

20 Числовые данные: две связанные группы

ПРОБЛЕМА

Мы имеем две выборки, которые соотносятся друг с другом, и одну интересующую нас числовую или ординальную переменную.

- Величина может быть измерена на каждом пациенте в двух состояниях. Например, в перекрёстном исследовании каждый пациент имеет два измерения на переменной: одно, когда принимает активное лечение, и одно, когда принимает плацебо.
- Пациенты в каждой выборке могут быть различными, но каким-то образом связанными друг с другом. Например, в исследовании «случай—контроль» пациенты одной группы могут быть индивидуально подобраны для пациентов в другой группе.

Такие данные известны как парные данные. Важно принять во внимание зависимость между двумя выборками при анализе данных, иначе преимущества парной связи будут потеряны. Мы делаем это, принимая во внимание разность в значениях для каждой пары, уменьшая таким образом наши две выборки до одной выборки разностей.

ПАРНЫЙ *t*-КРИТЕРИЙ

Допущения

В интересующей нас популяции индивидуальные разности нормально распределены с данной (обычно неизвестной) дисперсией. Мы имеем достаточный объём выборки, так что можем проверить допущение нормальности.

Обоснование

Если бы две группы измерений были одинаковыми, тогда мы ожидали бы среднюю разность между каждой парой измерений, равной нулю, в интересующей нас популяции, поэтому наша статистика критерия упрощается до одновыборочного t-критерия (см. главу 19) на разностях, где гипотетическое значение для средней разности в популяции равно нулю.

Дополнительное примечание

Из-за парной природы данных наши две выборки должны иметь одинаковый объём n. Мы имеем n разностей со средней выборочной $\overline{x}$ и оценкой стандартного отклонения s_d.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: средняя разница в популяции равна нулю.
  - ✧ H_1: средняя разница в популяции не равна нулю.
- 2. Отобрать необходимые данные из двух связанных выборок.
- 3. Рассчитать величину статистики критерия, отвечающей H_0, которая подчиняется t-распределению с $(n-1)$ степенями свободы:

$$t = \frac{(\overline{d} - 0)}{\mathrm{SE}(\overline{d})} = \frac{\overline{d}}{s_d/\sqrt{n}}.$$

- 4. Сравнить величину статистики критерия с величинами из известного распределения вероятности. Обратитесь к t-распределению Стьюдента в приложении А, табл. А2.
- 5. Интерпретировать значение достигнутого уровня значимости P и результаты.

Интерпретируйте P-значение и рассчитайте доверительный интервал для истинной средней разницы в популяции. 95% доверительный интервал задан посредством выражения:

$$\overline{d} \pm t_{0,05}\frac{s_d}{\sqrt{n}},$$

где $t_{0,05}$ — процентная точка t-распределения с $(n-1)$ степенями свободы, которая даёт двустороннюю вероятность в 0,05.

Если допущения не выполнены

Если разность не подчиняется нормальному распределению, допущение, лежащее в основе t-критерия, не удовлетворено. Мы можем или преобразовать данные, или применить непараметрический критерий, например, такой, как знаковый критерий, или критерий знаковых рангов Вилкоксона, чтобы оценить, сконцентрированы ли разности вокруг нуля.

КРИТЕРИЙ ЗНАКОВЫХ РАНГОВ ВИЛКОКСОНА

Обоснование

В главе 19 мы объяснили, как применять знаковый критерий на одной выборке числовых измерений для проверки нулевой гипотезы, когда медиана популяции равна определённому значению. Можно также использовать знаковый критерий, когда есть парные наблюдения, причём пары представляют собой подобранных пациентов (т.е. в исследовании «случай—контроль»), или измерения, проведённые на том же пациенте в различных условиях (как в перекрёстном исследовании двух видов лечения А и В). Для каждой пары оценивают разность в измерениях. Знаковый критерий можно применять для того, чтобы оценить, равна ли медиана разности в популяции нулю, рассматривая разности в выборке и отмечая, насколько она больше (или меньше) нуля. Однако знаковый критерий игнорирует информацию о величине этой разности.

Критерий знаковых рангов Вилкоксона принимает во внимание не только знаки разностей, но и их величину, и потому это более мощный критерий. Индивидуальную разность рассчитывают для каждой пары результатов. Игнорируя нулевые разности, их классифицируют как положительные или отрицательные. Кроме того, разности располагают в порядке их модуля, игнорируя их знаки, и соответственно ранжи-

руются. Наименьшая разность, таким образом, имеет ранг 1, вторая наименьшая — ранг 2 и т.д. вплоть до наибольшей разности, которая принимает ранг n', если имеется n' ненулевых разностей. Если две или более разности одинаковы, каждая получает среднее значение из рангов тех величин, которые были бы получены, если бы не были связаны. Под влиянием нулевой гипотезы об отсутствии различия сумма рангов, относящаяся к положительным и отрицательным разностям, должна быть одинаковой (см. следующий блок).

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: медиана разницы в популяции равна нулю.
  - ✧ H_1: медиана разницы в популяции не равна нулю.
- 2. Отобрать необходимые данные из двух взаимосвязанных выборок.
- 3. Вычислить величину статистики критерия, отвечающую H_0. Вычислите разность для каждой пары результатов. Проранжируйте все n' ненулевые разности, присваивая ранг 1 наименьшей разности и ранг n' наибольшей. Просуммируйте ранги положительных (T_+) и отрицательных (T_-) разностей.
  - ✧ Если $n' \leq 25$, статистика критерия T принимает значение T_+ или T_- в зависимости от того, какая из них меньше.
  - ✧ Если $n' > 25$, рассчитайте статистику критерия z, где:

$$z = \frac{\left|T - \frac{n'(n'+1)}{4}\right| - \frac{1}{2}}{\left(\frac{n'(n'+1)(2n'+1)}{24}\right)},$$

которая подчиняется нормальному распределению (её величина должна быть скорректирована, если имеется много связанных значений[1]).
- 4. Сравнить величину статистики критерия с величинами из известного распределения вероятности.
  - ✧ Если $n' \leq 25$, обратиться к распределению T в приложении А, табл. А8.
  - ✧ Если $n > 25$, обратиться к распределению z в приложении А, табл. А1.
- 5. Интерпретировать значение достигнутого уровня значимости P и результаты. Интерпретируйте значение P и рассчитайте доверительный интервал для медианы разностей по всей выборке.

ПРИМЕРЫ

96 новых призывников, все мужчины в возрасте 16–20 лет, прошли осмотр зубов, когда они зачислялись в Королевские Вооружённые Силы. После необходимого лечения их зубы были осмотрены через год. У человека 28 зубов, кроме зубов мудрости, и в этом исследовании каждый зуб имел 4 интересующие дантистов точки: каждый участник имел минимум 84 и максимум 112 измеряемых точек. Было интересно изучить эффект лечения на глубину дёсневого кармана (большая глубина кармана означает ухудшение состояния зуба).

Поскольку в этой выборке участников глубина кармана (принимая среднюю по измеряемым точкам во рту) была приблизительно нормально распределена, был использован парный t-тест для определения, была ли средняя глубина кармана такой же до и после лечения. Полный компьютерный результат анализа показан в приложении В.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: средняя разность усреднённой глубины дёсневого кармана человека до и после лечения в популяции призывников равна нулю.
  - ✧ H_1: средняя разность усреднённой глубины дёсневого кармана человека до и после лечения в популяции призывников не равна нулю.
- 2. Объём выборки n=96. Средняя разница усреднённой глубины дёсневого кармана $\bar{x}$=0,1486 мм. Стандартное отклонение разности s_d=0,5601 мм.
- 3. Статистика критерия:

$$t = \frac{0,1486}{0,5601/\sqrt{96}} = 2,60.$$

- 4. Мы обратимся к t-распределению в приложении А, табл. А2 с (96–1) степенями свободы: $0,01 < P < 0,05$ (результат компьютерного анализа даёт P=0,011).
- 5. Мы имеем данные за то, чтобы отбросить нулевую гипотезу, и можем заключить, что средняя глубина кармана у участников после лечения уменьшилась. 95% доверительный интервал для истинной средней разницы в усреднённой глубине кармана составляет 0,035–0,262 мм (т.е. $0,1486 \pm 1,95 \pm 0,5601/\sqrt{96}$). Конечно, нужно быть очень осмотрительными, если мы хотим сделать вывод, что именно лечение понизило среднюю глубину кармана, поскольку у нас не было контрольной группы, которая не получала лечения. Улучшение может быть следствием времени или изменения привычек зубной гигиены и может быть вызвано не лечением.

[1] Siegel S., Castellan N.J. Nonparametric Statistics for the Behavioural Sciences — 2nd edn — New York: McGraw–Hill, 1988.

Данные в нижеприведённой таблице показывают процент измеряемых сайтов, для которых была утрата прикрепления при каждой оценке в каждом из 14 этих участников. Потеря прикрепления — показатель заболевания дёсен, которое может быть более запущенным, чем оно оценено с помощью глубины кармана. Поскольку разница в процентах не была распределена нормально, мы использовали критерий знаковых рангов Вилкоксона для того, чтобы исследовать, имело ли лечение какое-либо влияние на утрату прикрепления.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: медиана разностей (до и после лечения) в процентах сайтов с потерей прикрепления равна нулю в популяции участников.
  - ✧ H_1: медиана разностей (до и после лечения) в процентах сайтов с потерей прикрепления не равна нулю в популяции участников.
- 2. Процент измеряемых сайтов с потерей прикрепления до и после лечения для каждого участника показан в таблице ниже.
- 3. Имеется одна нулевая разность: из остальных n'=13 разностей 3 положительны и 10 отрицательны. Сумма рангов положительных разностей T_+=3+5+13=21.
- 4. Поскольку n'<25, мы обращаемся к распределению T_+ в табл. А8 (приложение А): P >0,05 (результат компьютерного анализа P =0,09).
- 5. Имеется недостаточно данных, чтобы отбросить нулевую гипотезу об отсутствии изменения процента сайтов с потерей прикрепления. Медиана разностей в проценте сайтов с потерей прикрепления −3,1% (т.е. среднее от −2,5 до −3,6%), отрицательная медиана разностей означает, что в среднем процент сайтов с потерей прикрепления больше после лечения, хотя эта разница статистически не значима. Приложение А, табл. А7 показывает, что приближённый 95% доверительный интервал для медианы разностей в популяции дан 3-й и 12-й ранжированными разностями (включая нулевую): это −21,8 и 0,9%. Хотя результат критерия статистически незначим, более низкий предел показывает, что процент сайтов с потерей прикрепления мог быть на целых 21,8% больше после получения призывниками лечения.

Таблица 20-1. Исследование эффекта лечения на глубину дёсневого кармана

Призывник	1	2	3	4	5	6	7	8	9	10	11	12	13	14
До, %	65,5	75,0	87,2	97,1	100,0	92,6	82,3	90,0	93,0	100,0	91,7	97,7	79,0	95,4
После, %	100,0	10,0	100,0	97,1	99,1	100,0	91,6	94,6	95,5	97,3	92,3	98,0	100,0	99,0
Разность, %	−34,5	65,0	−12,8	0,0	0,9	−7,4	−9,3	−4,6	−2,5	2,7	−0,6	−0,3	−21,0	−3,6
Ранг	12	13	10	–	3	8	9	7	4	5	2	1	11	6

Duffy S. Results of a three year longitudinal study of early periodontitis in a group of British male adolescents // MSc Dissertation, University of London, Eastman Dental Institute for Oral Health Care Sciences, 1997.

21 Числовые данные: две независимые группы

ПРОБЛЕМА

Мы имеем выборки из двух независимых (несвязанных) групп пациентов и одну интересующую нас числовую или ординальную переменную. Интересуемся, одинаковые ли средние или распределение переменной в этих двух группах. Например, мы хотим сравнить вес в двух группах детей, причём каждому ребёнку назначена диета или плацебо случайным образом.

НЕПАРНЫЙ (ДВУХВЫБОРОЧНЫЙ) *t*-КРИТЕРИЙ

Допущения

В популяции переменная нормально распределена и дисперсии обеих групп одинаковы. Кроме того, размер выборки достаточный, чтобы проверить допущения нормальности и равных дисперсий.

Обоснование

Мы рассматриваем разность средних двух групп. При нулевой гипотезе, что популяционные (генеральные) средние в двух группах одинаковы, эта разность будет равна нулю, поэтому мы применяем статистику критерия, которая основана на разности в средних двух выборок и на величине разности популяционных средних при нулевой гипотезе (т.е. нулю). Эта статистика критерия, часто называемая t, подчиняется t-распределению Стьюдента.

Примечание

Наши две выборки имеют объёмы n_1 и n_2. Их средние — $\bar{x}_1$ и $\bar{x}_2$, их стандартные отклонения — s_1 и s_2.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: средние в двух популяциях равны.
  - H_1: средние в двух популяциях не равны.
- 2. Отобрать необходимые данные из двух взаимосвязанных выборок.
- 3. Вычислить величину статистики критерия, отвечающую H_0. Если s — объединённая оценка стандартного отклонения двух групп, то статистика критерия задана значением t, которая подчиняется t-распределению Стьюдента с $(n_1 + n_2 - 2)$ степенями свободы.

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}},$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\mathrm{SE}(\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

- 4. Сравнить величину статистики t-критерия с величинами из известного распределения вероятности. Обратитесь к t-распределению в приложении A, табл. A2. Когда объёмы выборки в двух группах большие, t-распределение аппроксимирует нормальное распределение, и тогда мы отбрасываем нулевую гипотезу при 5% уровне, если абсолютная величина (т.е. игнорируя знак) t более 1,96.
- 5. Интерпретировать величину P и результаты. Интерпретируйте величину P и рассчитайте доверительный интервал для разности этих двух средних. 95% доверительный интервал для этой разности задан выражением:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,05} \times \mathrm{SE}(\bar{x}_1 - \bar{x}_2),$$

где $t_{0,05}$ — процентная точка t-распределения с $(n_1 + n_2 - 2)$ степенями свободы, что даёт двустороннюю вероятность в 0,05.

Интерпретация доверительного интервала

Верхний и нижний пределы доверительного интервала могут применяться для того, чтобы оценить, важна ли клинически разница между двумя средними величинами. Например, если нижний предел близок к нулю, это означает, что истинная разница может быть очень маленькой и клинически не значимой, даже если критерий статистически значим.

Если допущения не удовлетворены[1]

Когда объёмы выборки достаточно большие, t-критерий в известной степени устойчив (робастный) к отклонениям от нормальности. Однако, он менее устойчив при неравных дисперсиях. Имеется модификация непарного t-критерия, которая разрешает неравные дисперсии, и её результаты часто даются в результатах компьютерного анализа. Однако, если вы сомневаетесь, что допущения удовлетворены, вы можете или преобразовать данные для достижения приближённой нормальности и/или равных дисперсий, или применить непараметрический критерий, например, такой, как двухвыборочный критерий Вилкоксона ранговых сумм.

ДВУХВЫБОРОЧНЫЙ КРИТЕРИЙ ВИЛКОКСОНА РАНГОВЫХ СУММ

Обоснование

Двухвыборочный критерий Вилкоксона ранговых сумм не содержит никаких предположений относительно распределения, это непараметрический аналог непарного t-критерия. Критерий основывается на сумме рангов величин в каждой из двух групп; последние должны быть сопоставимы после учёта различия в выборках, т.е. группы должны иметь подобное распределение. Эквивалентный критерий, известный как критерий Манна–Уитни, даёт идентичные результаты, хотя его сложнее выполнять вручную.

[1] С последствиями игнорирования ограничений на использование критерия Стьюдента читатели могут познакомиться в статье [84]. Прим. переводчика.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: две группы имеют одинаковое распределение в популяции.
  - ✧ H_1: две группы имеют различные распределения в популяции.
- 2. Отобрать необходимые данные из двух взаимосвязанных выборок.
- 3. Вычислить величину статистики критерия, отвечающую H_0. Все наблюдения ранжируют так, как если бы они были отобраны из одной выборки. Наблюдениям с равными значениями дают средний рангов, исходя из рангов, которые были бы получены, если бы не были обусловлены таким равенством. Затем сумму рангов T вычисляют в группе с меньшим объёмом выборки.
  - ✧ Если размер выборки в каждой группе 15 или меньше, T — статистика критерия.
  - ✧ Если по крайней мере одна из групп имеет размер выборки более 15, рассчитайте статистику критерия:

$$z = \frac{(T - \mu_T)}{\sigma_T},$$

которая подчиняется нормальному распределению, где:

$$\mu_T = \frac{n_S(n_S + n_L + 1)}{2} \qquad \sigma_T = \sqrt{\frac{n_L \mu_T}{6}},$$

и n_S и n_L — объёмы выборок соответственно наименьшей и наибольшей групп. z должно быть скорректировано, если имеется много связанных величин[1].

- 4. Сравнить значение статистики критерия с величинами из известного распределения вероятности. Если размер выборки в каждой группе 15 или меньше, обратитесь к распределению T в табл. А9 (приложение А). Если по крайней мере одна из групп имеет объём выборки более 5, обратитесь к z-распределению в приложении А, табл. А1.
- 5. Интерпретировать значение достигнутого уровня значимости P и результаты. Интерпретируйте значение P и получите доверительный интервал для разности двух медиан. При расчёте вручную это занимает много времени, так что детали не включены; некоторые статистические программы дадут доверительный интервал. Если этот расчёт не включён в ваш пакет программ, вы можете сослаться на доверительный интервал для медиан в каждой из двух групп.

ПРИМЕР 1

Чтобы определить эффект регулярной профилактики вдыхания кортикостероидов на эпизоды стридора, связанные с вирусной инфекцией у детей школьного возраста, было проведено рандомизированное двойное слепое контролируемое исследование, сравнивающее ингаляции беклометазона дипропионата с плацебо. В этом исследовании первичной конечной точкой был средний FEV1 за шестимесячный период. После проверки допущений нормальности и равной дисперсии (см. рис. 4-2) мы использовали непарный t-критерий для сравнений средних в двух группах. Полный листинг компьютерного результата анализа показан в приложении В.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: средние значения FEV1 в популяции детей школьного возраста в двух группах лечения одинаковы.
  - ✧ H_1: средние значения FEV1 в популяции детей школьного возраста в двух группах лечения не одинаковы.
- 2. Пролеченная группа: объём выборки n_1=50; среднее $\bar{x}_1$=1,64 л; стандартное отклонение s_1=0,29 л.

  Группа плацебо: объём выборки n_2=48; среднее $\bar{x}_2$=1,54 л; стандартное отклонение s_2=0,25 л.
- 3. Обобщённое стандартное отклонение:

$$s = \sqrt{\frac{(49 \times 0,29^2) + (47 \times 0,25^2)}{(50 + 48 - 2)}} = 0,2670 \text{ л.}$$

Статистика критерия

$$t = \frac{1,64 - 1,54}{0,2670 \times \sqrt{\frac{1}{50} + \frac{1}{48}}} = 1,9145.$$

- 4. Обратимся к t-распределению в приложении А, табл. А2 с 50+48−2=96 степенями свободы. Поскольку таблица А2 ограничена определёнными значениями степеней свободы, мы вынуждены интерполировать (определить требуемую величину, которая лежит между двумя известными величинами), поэтому мы интерполируем между величинами, относящимися к 50 и 100 степеням свободы. Следовательно, P >0,05 (результат компьютерного анализа даёт P =0,06).
- 5. У нас недостаточно данных для того, чтобы отбросить нулевую гипотезу при 5% уровне значимости. Однако, поскольку величина P только немногим больше 0,05, это может быть указанием, что средние двух популяций различны. Оценка разности между этими двумя средними 1,64−1,54=0,1 л. 95% доверительный интервал для истинного различия в двух средних варьируется от −0,006 до 0,206 л.

$$\left[= 0,10 \pm \left(1,96 \times 0,2670 \times \sqrt{\frac{1}{50} + \frac{1}{48}} \right) \right].$$

Данные любезно предоставлены Dr I. Doull, Cystic Fibrosis / Respiratory Unit, Department of Child Health, University Hospital of Wales, Cardiff, UK and Dr. F.C. Lampe, Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK.

[1] Siegel S., Castellan N.J. Nonparametric Statistics for the Behavioural Sciences. — 2nd edn. — New York: McGraw–Hill, 1988.

ПРИМЕР 2

Чтобы изучить, отличаются ли механизмы, включённые в фатальную соевую астму, от механизмов нормальной фатальной астмы, число CD3+ Т-клеток в подслизистой основе, показателе иммунной системы организма, сравнили с числом клеток в 7 случаях фатальной соевой астмы, вызванной пылью, и в 10 случаях фатальной астмы. Из-за маленьких размеров выборки и явно скошенных (асимметричных) данных мы использовали двухвыборочный критерий Вилкоксона ранговых сумм для сравнения распределений.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: распределения количества CD3+ Т-клеток в двух популяциях одинаковы.
  - H_1 распределения количества CD3+ Т-клеток в двух популяциях не одинаковы.
- 2. Группа соевой астмы: объём выборки n_s=7, уровни CD3+ Т-клеток (клетки/мм^2) были 34,45; 0,00; 1,36; 0,00; 1,43; 0,00; 4,01.
  Группа астмы: объём выборки n_L=10, уровни CD3+ Т-клеток (клетки/мм^2) были 74,17; 13,75; 37,50; 1225,51; 99,99; 3,76; 58,33; 73,63; 4,32; 154,86. Проранжированные данные показаны в таблице ниже.
- 3. Сумма рангов в группе астмы равна 2+2+2+4+5+ +7+10=32. Сумма рангов в группе астмы равна 6+8+9+11+12+13+14+15+16+17=121.
- 4. Поскольку в каждой группе 10 или меньше значений, мы получаем величину P из табл. А9 (приложение А): P <0,01 (результат компьютерного анализа даёт P=0,002).
- 5. Имеются данные, чтобы отбросить нулевую гипотезу, что распределения уровней CD3+ Т-клеток в двух группах одинаковы. Среднее количество CD3+ Т-клеток в группах с соевой и фатальной астмой составляет 1,36 (95% доверительный интервал от 0 до 34,45) и (58,33+73,63)/2=65,98 (95% доверительный интервал соответственно от 4,32 до 154,86) клеток/мм^2. Мы, таким образом, верим, что число клетки CD3+ Т-клеток понижено в фатальной соевой астме, и приходим к мысли, что в большинстве смертей от такой астмы задействован другой механизм, а не описанный.

Таблица 21-1. Изучение механизмов фатальной соевой астмы

Соевая астма	**0,0**	**0,0**	**0,0**	**1,36**	**1,43**		**4,01**			**34,45**							
Астма						3,76		4,32	13,75		37,5	58,33	73,63	74,17	99,99	154,86	1225,51
Ранг	2	2	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Данные любезно предоставлены Dr M. Synek, Coldeast Hospital, Sarisbury, and Ms F.C. Lampe, Department of Primary Care and Population Sciences, Royal Free and University College School of Medicine, Royal Free Campus, London UK.

22 Числовые данные: более двух групп

ПРОБЛЕМА

У нас есть выборки из ряда независимых групп. Мы имеем одну числовую или ординальную переменную и интересуемся, варьируют ли величины переменной среди групп, например, варьирует ли количество тромбоцитов у женщин различного этнического происхождения. Можно провести сравнения значений для каждой из возможных пар групп, но высокое значение ошибок 1-го рода в результате большого числа сравнений означает, что можно сделать некорректные заключения, поэтому используем один общий критерий для определения, имеются ли различия между всеми группами сразу.

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Допущения

Группы определены с помощью уровней одного фактора (т.е. отличающихся условиями происхождения). В интересующей нас популяции переменная нормально распределена (в каждой группе!), и дисперсия во всех группах одинакова. Объём выборки достаточный, чтобы проверить эти допущения.

Обоснование

Однофакторный дисперсионный анализ включает общую вариацию в данных на вариацию, которую можно отнести к отличиям между пациентами из разных групп (межгрупповая дисперсия), и случайную вариацию между пациентами внутри каждой группы (внутригрупповая дисперсия, иногда называется необъяснённая, или вариация остатков). Эти компоненты вариации измеряют с применением дисперсий, отсюда и название дисперсионный анализ (ANOVA — ANalysis Of VAriance). При нулевой гипотезе, что групповые средние одинаковы, межгрупповая дисперсия будет подобна внутригрупповой. Если, однако, между группами есть различия, то межгрупповая дисперсия будет больше, чем внутригрупповая. Критерий основан на соотношении двух этих дисперсий.

Примечание

Есть k независимых выборок, причём каждая определяет отдельную группу. Объёмы выборок, средние и стандартные отклонения в каждой группе соответственно n_i, $\bar{x}_i$ и s_i (i=1, 2, .., k). Общий объём выборок $n=n_1+ n_2+...+n_k$.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: все групповые средние в популяции равны.
  - H_1: по крайней мере одна групповая средняя в популяции отличается от других.
- 2. Отобрать необходимые данные из двух взаимосвязанных выборок.
- 3. Вычислить величину статистики критерия, отвечающую H_0. Статистика критерия для ANOVA — F, отношение межгрупповой дисперсии к внутригрупповой. Эта F-статистика подчиняется F-распределению Фишера–Снедекора с (k–1, n–k) степенями свободы соответственно в числителе и знаменателе. Расчёты, содержащиеся в ANOVA, сложны и здесь не показаны. Большинство компьютерных программ выдают величины прямо в таблице ANOVA, которая обычно включает F-отношение и величину P (см. пример 1).
- 4. Сравнить значение статистики F-критерия со значением из известного распределения вероятности. Обратимся к F-распределению в приложении А, табл. А5. Поскольку межгрупповая дисперсия ≥ внутригрупповой, мы смотрим на односторонние значения P.
- 5. Интерпретировать величину P и результаты. Если получен статистически значимый результат на этой начальной стадии, можно рассматривать проведение специальных парных апостериорных сравнений. Мы можем применить один из ряда специальных критериев, разработанных для этой цели (например, Дункана, Шеффе), или непарный t-критерий, скорректированный для проверки множественных гипотез. Можно также рассчитать доверительный интервал для средней в каждой отдельной группе. Отметьте, что мы применяем обобщённую оценку дисперсии значений от всех групп при расчёте доверительных интервалов и использовании t-критериев. Большинство программ называет эту оценку дисперсии дисперсией остатков, или средним квадратом остатков; её можно найти в таблице ANOVA.

Хотя эти два критерия, по-видимому, различны, непарный t-критерий и ANOVA дают эквивалентные результаты, когда имеется только две группы пациентов.

Если допущения не удовлетворены

Хотя ANOVA относительно устойчив при небольших отклонениях от нормальности, он не устойчив при неравных дисперсиях, поэтому перед выполнением анализа проверяем нормальность и тестируем, равны ли генеральные дисперсии в группах, «на глазок» или применяя тесты Левене или Бартлетта. Если допущения не удовлетворены, можно преобразовать данные или применить непараметрический аналог однофакторного ANOVA, критерий Крускала–Уоллиса.

КРИТЕРИЙ КРУСКЕЛА-УОЛЛИСА

Обоснование

Этот непараметрический критерий — расширение двухвыборочного критерия Вилкоксона ранговых сумм. При нулевой гипотезе отсутствия различий в распределениях между группами суммы рангов в каждой из k групп должны быть сравнимы после учёта любых различий в размере выборки.

- **1. Определить нулевую и альтернативную изучаемые гипотезы.**
  - ✧ H_0: каждая группа имеет одинаковое распределение величин в популяции.
  - ✧ H_1: каждая группа не имеет одинаковое распределение величин в популяции.
- **2. Отобрать необходимые данные из двух взаимосвязанных выборок.**
- **3. Вычислить величину статистики критерия, отвечающую H_0.**

Проранжируйте все n значений и рассчитайте сумму рангов в каждой из групп: эти суммы — $R_1, ..., R_k$. Статистика критерия (которая должна быть модифицирована, если имеется много связанных значений[1]), даётся посредством формулы:

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1),$$

которая подчиняется распределению Хи-квадрат Пирсона с $(k-1)$ степенями свободы.

- **4. Сравнить величину критерия статистики с величинами из известного распределения вероятности.** Обратимся к распределению Н в приложении А, табл. А3.
- **5. Интерпретировать величину P и результаты.** Интерпретируйте величину P и, если результат статистически значим, используйте двухвыборочные непараметрические критерии, корректируя для множественного тестирования. Рассчитайте доверительный интервал для медианы в каждой группе.

Мы применяем однофакторный ANOVA или его непараметрический эквивалент, когда группы соотносятся с одним фактором и независимы. Можно использовать другие виды ANOVA, когда план исследования более сложен[2].

ПРИМЕР 1

Всего 150 женщин различного этнического происхождения были включены в перекрёстное исследование факторов, относящихся к свёртыванию крови. Мы сравнили средние уровни тромбоцитов в четырёх группах, используя однофакторный ANOVA. Допущения (нормальность, постоянная вариация) были удовлетворены, как показано в результатах компьютерного анализа (приложение В).

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: нет разницы в средних уровнях тромбоцитов в четырёх группах в популяции.
  - ✧ H_1: среднее количество тромбоцитов по крайней мере в одной группе отличается от других в популяции.
- 2. Табл. 22-1 суммирует данные в каждой группе.

Таблица 22-1. Исследование факторов свёртывания крови у женщин различных национальностей

Группа	Объём выборки n %	Среднее (×10⁹) $\overline{x}$	Стандартное отклонение (×10⁹), s	95% ДИ (использовано обобщённое стандартное отклонение, см. п. 3)
Жительницы Кавказа	90 (60,0)	268,1	77,08	От 252,7 до 283,5
Жительницы Карибского бассейна	21 (14,0)	254,3	67,50	От 220,9 до 287,7
Жительницы Средиземноморья	19 (12,7)	281,1	71,09	От 245,7 до 316,5
Другие	20 (13,3)	273,3	63,42	От 238,9 до 307,7

- 3. Табл. 22-2 взята из результатов компьютерного анализа.

Таблица 22-2. Исследование факторов свёртывания крови у женщин различных национальностей (ANOVA)

Источник вариации	Сумма квадратов	Степени свободы	Средний квадрат	F-отношение	Достигнутый уровень значимости P
Межгрупповая дисперсия	7711,967	3	2570,656	0,477	0,6990
Внутригрупповая дисперсия	787289,533	146	5392,394		

[1] Siegel S., Castellan N.J. Nonparamelric Statistics for the Behavioral Sciences. New York: McGraw–Hill, 1988.

[2] Mickey R.M., Dunn O.J., Clark, V.A. Applied Statistics: Analysis of Variance and Regression. — 3rd edn — Chichester: Wiley, 2004.

Обобщённое стандартное отклонение = $= \sqrt{5392,394} \times 10^9 = 73,43 \times 10^9$.

- 4. Таблица ANOVA даёт P =0,7. (Мы могли обратиться к F-распределению в приложении А, табл. А5 с 3,146 степенями свободы для определения величины P.)
- 5. Недостаточно данных, чтобы отбросить нулевую гипотезу, что средние уровни в четырёх группах в популяции одинаковы.

Данные любезно предоставлены Dr. R.A. Kadir, University Department of Obstetrics and Gynaeclogy, and Professor C.A. Lee, Heemophilia Centre and Haemostatsis Unit, Royal Free Hospital, London, UK.

ПРИМЕР 2

Уровень качества жизни, измеренный с применением анкеты SF-36, был получен от трёх групп пациентов: с тяжёлой гемофилией, с лёгкой/умеренной гемофилией и от контроля. Каждая группа состояла из 20 пациентов. Уровень шкалы физической активности (ШФА), который мог принимать значения от 0 до 100, сравнили в этих трёх группах. Поскольку визуальное рассмотрение на рис. 22-1 показало, что данные не были нормально распределены, мы использовали критерий Крускела–Уоллиса.

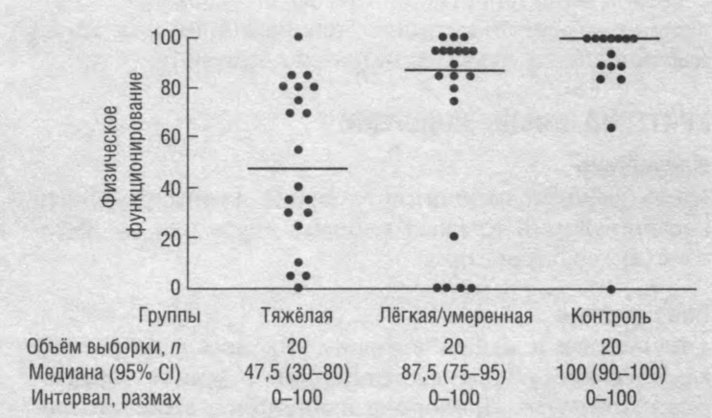


Рис. 22-1. Точечный график, показывающий уровни физического функционирования (из анкеты SF-36) у пациентов с тяжёлой и лёгкой/умеренной гемофилией и у контроля. Горизонтальные линии — медианы.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: каждая группа в популяции имеет одинаковое распределение шкалы ШФА.
  - H_1: по крайней мере одна из групп в популяции имеет другое распределение шкалы ШФА.
- 2. Данные показаны на рис. 22-1.
- 3. Сумма рангов в группе тяжёлой гемофилии =372. Сумма рангов в группе умеренной гемофилии =599. Сумма рангов в контрольной группе =859:

$$H = \frac{12}{60(60+1)}\left(\frac{372^2}{20} + \frac{599^2}{20} + \frac{859^2}{20}\right) - 3(60+1) = 19,47.$$

- 4. Обратимся к распределению H в приложении А, табл. А3: P <0,001.
- 5. Есть убедительные данные для отбрасывания нулевой гипотезы, что распределение уровня ШФА в трёх группах одинаково. Провели парные сравнения с применением двухвыборочных критериев Вилкоксона ранговых сумм, корректируя величины P для числа проведённых тестов. Пациенты как с тяжёлой, так и с умеренной гемофилией имели значительно более низкий уровень ШФА, чем контроль (P =0,0003 и P=0,03 соответственно), но распределения уровня ШФА в группах с гемофилией незначительно отличались друг от друга (P =0,09).

Данные любезно предоставлены Dr A. Miners, Department of Primary Care and Population Sciences, Royal Free and University College Medical School, Royal Free Campus, London, UK, и Dr C. Jenkinson, Health Services Research Unit, University of Oxford, Oxford, UK.

23 Качественные данные: одна пропорция

ПРОБЛЕМА

Мы имеем единичную выборку из n пациентов; каждый пациент обладает интересующей нас характеристикой (например, мужчина, беременна, умер/ла) или не обладает такой характеристикой [например, женщина, не беременна, жив(а)]. Полезная сводка данных обеспечивается значением пропорции, доли пациентов с данной характеристикой. Мы заинтересованы в определении, имеет ли истинная пропорция интересующей нас популяции некую конкретную величину.

КРИТЕРИЙ ОДНОЙ ПРОПОРЦИИ

Допущения

Наша выборка пациентов отобрана из интересующей нас популяции. Каждый пациент имеет или не имеет особую характеристику.

Обозначения

r пациентов в нашей выборке объёмом n имеют эту характеристику. Оценка пропорции с данной характеристикой $p=r/n$. Пропорция пациентов с этой характеристикой в популяции — π. Мы заинтересованы определить, равно ли π некоторой конкретной величине π_1.

ОБОСНОВАНИЕ

Число пациентов с данной характеристикой подчиняется биномиальному распределению, но его можно аппроксимировать нормальным распределением при условии, что каждая из величин np и $n(1-p)$ более 5. В этом случае p приблизительно нормально распределяется с оценкой среднего значения, равным p и оценкой стандартного отклонения.

$$\sqrt{\frac{p(1-p)}{n}}.$$

Таким образом, наша статистика критерия, которая основывается на p, также имеет нормальное распределение.

- **1. Определить нулевую и альтернативную исследуемые гипотезы.**
  - H_0: пропорция популяции π равна некоторой величине π_1.
  - H_1: популяционная пропорция π не равна π_1.
- **2. Отобрать необходимые данные из выборки пациентов.**
- **3. Рассчитать величину z-статистики, отвечающей H_0:**

$$z=\frac{|p-\pi_1|-\frac{1}{2n}}{\sqrt{\frac{p(1-p)}{n}}},$$

которая подчиняется нормальному распределению. Величина $^1/_2n$ в числителе — поправка Йэйтса на непрерывность; она включена, чтобы учесть, что мы аппроксимируем дискретное биномиальное распределение с помощью непрерывного нормального распределения.

- **4. Сравнить величину тестовой статистики с величинами из известного распределения вероятности.** Обратитесь к распределению z в приложении А, табл. А1.
- **5. Интерпретировать значение уровня значимости P и результаты.** Интерпретировать величину P и рассчитать доверительный интервал для популяционной пропорции π. 95% доверительный интервал для π равен:

$$p \pm 1{,}96\sqrt{\frac{p(1-p)}{n}}.$$

Мы можем использовать этот доверительный интервал для оценки клинической или биологической значимости результатов. Широкий доверительный интервал — указание на то, что наша оценка имеет плохую точность.

ЗНАКОВЫЙ КРИТЕРИЙ В ПРИЛОЖЕНИИ К ПРОПОРЦИИ

Обоснование

Знаковый критерий можно применять, если интересующий ответ можно выразить как предпочтение (например, в перекрёстном исследовании пациенты могут выбрать лечение А или В). Если выбора в общем нет, мы должны ожидать пропорцию, предпочитающую А, скажем, равную $^1/_2$. Мы применяем знаковый критерий для того, чтобы оценить, так ли это.

Хотя эта формулировка проблемы и статистика её критерия, по-видимому, отличается от описанных в главе 19, оба подхода к статистике критерия дают одинаковый результат.

- **1. Определить нулевую и альтернативную исследуемые гипотезы.**
  - H_0: пропорция π выбора для А в популяции равна $^1/_2$.
  - H_1: пропорция выбора для А в популяции не равна $^1/_2$.
- **2. Отобрать необходимые данные из выборки пациентов.**
- **3. Рассчитать значение статистики критерия, отвечающей H_0.** Не принимайте во внимание пациентов, которые не имеют выбора и соответственно уменьшают размер выборки с n до n'. Тогда $p=r/n'$, где r — число выборов для А.
  - Если $n' \leq 10$, сосчитайте r — число выборов для А.
  - Если $n' > 10$, рассчитайте статистику критерия:

$$z' = \frac{\left|p - \frac{1}{2}\right| - \frac{1}{2n'}}{\sqrt{\frac{p(1-p)}{n'}}},$$

z' имеет нормальное распределение. Отметьте, что эта формула основана на статистике критерия z, которую мы применяли в предыдущей главе для проверки нулевой гипотезы, что пропорция в популяции равна π_1; здесь мы заменяем n на n', а π_1 — на $^1/_2$.

- **4. Сравните величину статистики z-критерия с величинами из известного распределения вероятности.**
  - ✧ Если $n' \leq 10$, используйте r из приложения А, табл. А6.
  - ✧ Если $n' > 10$, используйте z' из приложения А, табл. А1.
- **5. Интерпретировать величину P и результаты.** Интерпретируйте величину P и вычислите доверительный интервал для пропорции выбора А во всей выборке размером n.

ПРИМЕР 1

Вирус герпеса человека (HHV-8) связан с саркомой Капоши, первичной лимфомой и отдельными типами многоцентрового заболевания Кастлемана. Предполагается, что HHV-8 может передаваться половым путём. Чтобы оценить соотношения между сексуальным поведением и инфекцией HHV-8, распространённость антител к HHV-8 была определена в группе из 271 гомо/бисексуального мужчины, посетивших Лондонскую клинику заболеваний, передаваемых половым путём. В сыворотке крови донорской популяции в Англии серораспространённость HHV-8 составляет 2,7%. Вначале распространённость от этого исследования сравнивалась с 2,7% с применением теста одной пропорции.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: серораспространённость HHV-8 в популяции гомо/бисексуальных мужчин равна 2,7%.
  - ✧ H_1: серораспространённость HHV-8 в популяции гомо/бисексуальных мужчин не равна 2,7%.
- 2. Объём выборки n=271; число пациентов, серopоложительных к HHV-8, r=50. Серораспространённость p=50/271=0,185 (т.е. 18,5%).
- 3. Статистика z-критерия равна:

$$z = \frac{|0{,}185 - 0{,}027| - \frac{1}{2 \times 271}}{\sqrt{\frac{0{,}185 \times (1 - 0{,}185)}{271}}} = 6{,}62.$$

- 4. Обратимся к распределению z в приложении А, табл. А1: $P < 0{,}0001$.
- 5. Имеются надёжные данные о том, что серораспространённость HHV-8 у гомо/бисексуальных мужчин, посещающих данную клинику в Англии, выше, чем в крови донорской популяции. 95% доверительный интервал для серораспространённости HHV-8 в популяции гомо/бисексуальных мужчин составляет от 13,9 до 23,1% и рассчитывается так:

$$\left\{0{,}185 \pm 1{,}96 \times \sqrt{\frac{0{,}185 \times (1 - 0{,}185)}{271}}\right\} \times 100\%.$$

Данные любезно предоставлены N.A. Smith, D. Barlow, and B.S.Peters, Department of Genitourinary Medicine, Guy's and St Thomas' NHS Trust, London, и J. Best, Department of Virology, Guy's, King's College and St Thomas's School of Medicine, King's College, London, UK.

ПРИМЕР 2

В двойном слепом перекрёстном исследовании 36 взрослых с аллергическим ринитом лечились подкожными инъекциями или ингаляцией аллергенов или плацебо, причём каждое лечение проводилось ежедневно в течение определённого периода. Пациентов спрашивали, предпочитают ли они активное лечение или плацебо. Знаковый критерий был использован для того, чтобы изучить, были ли одинаковыми пропорции пациентов, предпочитающих эти два препарата.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: пропорция пациентов в популяции, предпочитающих активный препарат, равна 0,5.
  - H_1: пропорция пациентов, предпочитающих активный препарат, не равна 0,5.
- 2. Из 36 взрослых 27 выразили предпочтение; 21 из них предпочли активный препарат. Из пациентов с выбором доля предпочитающих активный препарат p=21/27=0,778.
- 3. Статистика критерия:

$$z' = \frac{|0,778 - 0,5| - \frac{1}{2 \times 27}}{\sqrt{\frac{0,778 \times (1 - 0,778)}{27}}} = 3,24$$

- 4. Обратимся к описанию z' в приложении А, табл. А1: P=0,001.
- 5. Это существенное свидетельство, позволяющее отклонить нулевую гипотезу о том, что эти два препарата в популяции выбирают одинаково часто. 95% доверительный интервал для истинной пропорции, рассчитанный как: $0,778 \pm 1,96 \times \sqrt{\frac{0,778 \times (1 - 0,778)}{27}}$, составляет от 0,62 до 0,94.

Таким образом, мы полагаем, что по крайней мере почти две трети лиц в популяции предпочитают активный препарат.

24 Качественные данные: две пропорции

ПРОБЛЕМА[1]

- Есть две независимые группы пациентов (например, гомосексуальные мужчины с гонореей и без гонореи в анамнезе). Мы хотим узнать, одинаковы ли доли пациентов с определённым свойством (например, инфицированные вирусом герпеса человека HHV-8) в этих двух группах.
- Есть две связанные группы, т.е. пациенты могут быть специально подобранными или обследованными дважды в различных условиях (скажем, до и после лечения). Мы хотим узнать, одинаковы ли доли пациентов с определённым свойством в этих двух группах.

НЕЗАВИСИМЫЕ ГРУППЫ: КРИТЕРИЙ ХИ-КВАДРАТ ПИРСОНА

Терминология

Данные изначально получены как частоты, т.е. число пациентов с определённым свойством и без него в каждой выборке. Таблица, в которой исходные данные — частоты, называется таблицей сопряжённости; когда эта таблица имеет два ряда и две колонки, она называется таблицей 2×2. Табл. 24-1 показывает наблюдаемые частоты в этих четырёх клетках, соответствующие каждой комбинации ряд/колонка, четыре частные суммы (частота в определённом ряду или колонке, например, $a+b$), и общее число пациентов n. Мы можем рассчитать частоту, которую ожидаем в каждой из четырёх клеток таблицы, если гипотеза H_0 истинна (ожидаемые частоты).

Допущения

Есть выборки размеров n_1 и n_2 из двух независимых групп пациентов. Нас интересует, одинаковы ли доли пациентов, обладающих определённым свойством, в этих двух группах. Каждый пациент представлен в исследовании только один раз. Ряды (и колонки) таблицы взаимоисключающие: каждый пациент может находиться только в одном ряду и только в одной колонке. Обычный, хотя и консервативный, подход требует, чтобы ожидаемая частота в каждой из этих четырёх клеток была по крайней мере не меньше пяти.

Обоснование

Если пропорции лиц с определённым свойством в этих двух группах равны, мы можем оценить общую пропорцию таких пациентов с помощью выражения $p=(a+b)/n$; мы ожидаем, что их будет $n_1 \times p$ в группе 1 и $n_2 \times p$ — в группе 2. Подобным же образом мы оцениваем и ожидаемые значения для пациентов без данного свойства. Таким образом, каждая ожидаемая частота — произведение двух частных сумм фактических частот, делённых на общую сумму частот. Большое расхождение между наблюдаемыми (O) и соответственно ожидаемыми (E) частотами — показатель того, что пропорции в этих двух группах различаются. Статистика критерия основывается именно на этом расхождении.

Таблица 24-1. Наблюдаемые частоты

Свойство	Группа 1	Группа 2	Всего
Имеется	a	b	$a+b$
Отсутствует	c	d	$c+d$
Всего	$n_1=a+c$	$n_2=b+d$	$n=a+b+c+d$
Доля (пропорция) пациентов с определённым свойством	$p_1=a/n_1$	$p_2=b/n_2$	$p=(a+b)/n$

- **1. Определить нулевую и альтернативную изучаемые гипотезы.**
  - H_0: пропорции пациентов с данным свойством равны в этих двух популяциях.
  - H_1: эти популяционные пропорции не равны.
- **2. Отобрать необходимые данные из выборок лиц.**
- **3. Рассчитать величину статистики критерия, отвечающей H_0:**

$$\chi^2 = \sum \frac{\left(|O-E| - \frac{1}{2}\right)^2}{E},$$

где O (от observed — наблюдаемая, фактическая) и E (от expected — ожидаемая, теоретическая — перев.) — соответственно наблюдаемая и ожидаемая частоты в каждой из четырёх клеток таблицы. Вертикальные линии вокруг $O-E$ означают, что мы игнорируем знак разности (используем модуль числа). $^1/_2$ в числителе — поправка Йэйтса на непрерывность. Статистика этого критерия следует Хи-квадрат распределению Пирсона с одной степенью свободы.

- **4. Сравнить величину статистики критерия с величинами из известного распределения вероятности.** Обратиться к распределению Хи-квадрат в приложении А, табл. А3.
- **5. Интерпретировать величину P и результаты.** Интерпретируйте величину P (достигнутый уровень значимости) и рассчитайте доверительный интервал для разности популяционных пропорций. 95% доверительный интервал вычисляется по формуле:

$$(p_1 - p_2) \pm 1{,}96\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}.$$

[1] Несколько программ для анализа таких данных размещены по адресу http://www.biometrica.tomsk.ru/lib2.htm Для использования данных программ читателю следует лишь ввести свои частоты в клетки таблицы и получить результаты, содержащие все необходимые статистики. Прим. переводчика.

Если допущения не выполняются

Если $E<5$ в какой-либо одной из клеток, мы используем точный критерий Фишера для получения значения достигнутого уровня значимости P, который не основывается на аппроксимации распределения Пирсона Хи-квадрат. Это лучше всего выполнить с помощью компьютерной программы, поскольку расчёты утомительны для ручной работы.

СВЯЗАННЫЕ ГРУППЫ: КРИТЕРИЙ МАКНЕМАРА

Допущения

Две группы могут быть связанными или зависимыми, т.е. каждый пациент может быть обследован в двух различных состояниях, условиях. Каждый пациент классифицируется согласно тому, имеется ли определённое свойство в обоих состояниях, или только в одном состоянии, или же ни в одном из них (табл. 24-2).

Таблица 24-2. Наблюдаемые частоты пар, в которых есть характеристика или её нет

Состояние 2	Состояние 1		Общее количество пар
	Свойство имеется	Свойство отсутствует	
Свойство имеется	w	x	$w+x$
Свойство отсутствует	y	z	$y+z$
Общее количество пар	$w+y$	$x+z$	$m=w+x+y+z$

Обоснование

Наблюдаемые пропорции с наличием определённого свойства в этих двух состояниях — $(w+y)/m$ и $(w+x)/m$. Они будут различаться, если x и y различаются. Таким образом, для сравнения пропорций с наличием определённого свойства мы не принимаем во внимание тех пациентов, которые согласованы в этих двух состояниях, и сконцентрируемся на несогласованных парах x и y.

- **1. Определить нулевую и альтернативную исследуемые гипотезы**.
  - H_0: пропорции с наличием определённого свойства в двух популяциях одинаковы.
  - H_1: эти популяционные пропорции не одинаковы.
- **2. Отобрать необходимые данные из двух выборок.**
- **3. Рассчитать величину статистики критерия, отвечающей H_0:**

$$\chi^2 = \frac{\left(|x-y|-1\right)^2}{x+y},$$

подчиняющейся распределению Хи-квадрата с одной степенью свободы. Единица в числителе — поправка Йэйтса на непрерывность.
- **4. Сравнить величину статистики критерия с величинами из известного распределения вероятности.** Обратимся к Хи-квадрат распределению в приложении А, табл. А3.
- **5. Интерпретировать величину P и результаты.** Интерпретируйте величину достигнутого уровня значимости P и рассчитайте доверительный интервал для разницы в популяционных пропорциях. Приблизительный 95% ДИ вычисляется так:

$$\frac{x-y}{m} \pm \frac{1{,}96}{m}\sqrt{x+y-\frac{(x-y)^2}{m}}.$$

ПРИМЕР 1

Чтобы оценить соотношение между сексуальными факторами риска и инфекцией HHV-8 (исследование, описанное в главе 23), было проведено сравнение распространённости серopоложительности к HHV-8 у гомо/бисексуальных мужчин с гонореей в анамнезе и у мужчин, которые ранее не имели гонорею в анамнезе, с использованием критерия Хи-квадрат Пирсона. Типичная компьютерная распечатка (листинг) показана в приложении В.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: частота сероpаспространённости одинакова у мужчин популяции с гонореей и без гонореи в анамнезе.
  - H_1: частоты сероpаспространённости не одинаковы в этих двух популяциях.
- 2. Наблюдаемые частоты отображены в следующей таблице сопряжённости: 14/43 (32,6%) и 36/228 (15,8%) мужчин с гонореей и без неё в анамнезе серопозитивны по HHV-8 соответственно.
- 3. Ожидаемые частоты показаны в четырёх клетках таблицы сопряжённости.
  - Статистика критерия Хи-квадрат Пирсона равна:

$$\chi^2 = \left\{\frac{\left(|14-7{,}93|-\frac{1}{2}\right)^2}{7{,}93}+\frac{\left(|36-42{,}07|-\frac{1}{2}\right)^2}{42{,}07}+\frac{\left(|29-35{,}07|-\frac{1}{2}\right)^2}{35{,}07}+\frac{\left(|192-185{,}93|-\frac{1}{2}\right)^2}{185{,}93}\right\}=5{,}70.$$

- 4. Обратимся к Хи-квадрат распределению Пирсона в приложении А, табл. А3 с одной степенью свободы: $0{,}01 < P < 0{,}05$ (компьютерный результат даёт значение P=0,017).
- 5. Мы имеем убедительное подтверждение реального различия в частотах серораспространённости в двух популяциях. Мы оцениваем это различие как 32,6%−15,8%=16,8%. 95% ДИ для истинного различия в этих двух процентных отношениях составляет от 2 до 31,6% [т.е. 16,8±1,96× ×√({32,6×67,4}/43+{15,8×84,2}/228)].

Таблица 24-3. Соотношение между сексуальными факторами риска и инфекцией HHV-8

Реакция на HHV-8	Наличие гонореи в анамнезе: Да — Фактическая частота	Да — Ожидаемая частота	Нет — Фактическая частота	Нет — Ожидаемая частота	Всего
Серопозитивны	14	(43×50/271)=7,93	36	(228×50/271)=42,07	50
Серонегативны	29	(43×221/271)=35,07	192	(228×221/271)=185,93	221
Всего	43		228		271

ПРИМЕР 2

Чтобы сравнить два метода определения состояния зубов (есть или нет кариес), дантист, применяя радиографию, оценил состояние первых 100 больших коренных зубов, которые имели крошечные полости или не имели их вообще. Эти результаты сравнили с результатами, полученными с применением более объективного подхода визуальной оценки сектора каждого зуба. Процентные соотношения зубов, обнаруженных как имеющие полости с помощью этих двух методов, сравнили с применением критерия Макнемара.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: два метода оценки идентифицируют один и тот же процент зубов с полостями в популяции.
  - H_1: эти проценты не равны.
- 2. Частоты для согласованных пар изображены в табл. 24-4.

Таблица 24-4. Частоты для согласованных пар

Диагностика по секторам (визуальная)	Радиографическая диагностика: Полости обнаружены	Полости не обнаружены	Всего
Полости обнаружены	45	4	49
Полости не обнаружены	17	34	51
Всего	62	38	100

- 3. Статистика критерия Хи-квадрат Пирсона равна:

$$\chi^2 = \frac{(|17-4|-1)^2}{17+4} = 6{,}86.$$

- 4. Обратившись к табл. А3 распределения Пирсона (в приложении А), имеем для одной степени свободы: $0{,}001 < P < 0{,}01$ (компьютерный результат 0,009).

Имеются убедительные подтверждения для отказа от нулевой гипотезы о том, что один и тот же процент обнаружения полости в зубах обеспечивается применением этих двух методов исследования. Радиографический метод имеет тенденцию не обнаруживать полости. Мы оценили разницу в процентном соотношении зубов, обнаруженных как имеющие полости, как 51%−38%=13%. Приближённый доверительный интервал для истинной разницы в процентах задан от 4,4 до 21,6%, т.е.

$$\left\{\frac{|17-4|}{100} \pm \frac{1{,}96}{100} \times \sqrt{(17+4) - \frac{(17-4)^2}{100}}\right\} \times 100\%.$$

Данные взяты из: Ketley C.E., Holt R.D. Visual and radiographic diagnosis of occlusal caries in first permanent molars and in second primary molars. British Dental Journal. — 1993. — Vol. 174. — P. 364–370.

25 Качественные данные: более двух категорий

КРИТЕРИИ ХИ-КВАДРАТ ПИРСОНА: БОЛЬШИЕ ТАБЛИЦЫ СОПРЯЖЁННОСТИ

Проблема

Пациентов можно классифицировать по двум факторам. Например, один фактор может представлять собой тяжесть заболевания (лёгкое, средней тяжести или тяжёлое), а другой фактор — группу крови (A, B, O, AB). Нас интересует, взаимосвязаны ли эти два фактора. Возможно, лица с определённой группой крови имеют большую тяжесть заболевания?

Допущения

Данные могут быть представлены в $r \times c$ таблице сопряжённости с r рядами и c колонками (табл. 25-1). В таблицу вводятся частоты; каждая клетка содержит число лиц в отдельном ряду и отдельной колонке. Каждый пациент представлен один раз и может находиться только в одном ряду и одной колонке, т.е. категории (уровни, градации) каждого фактора альтернативные, взаимоисключающие. По крайней мере 80% ожидаемых частот больше или равны 5.

Обоснование

Нулевая гипотеза состоит в том, что между этими двумя факторами связи нет. Отметьте, что если есть только два ряда и две колонки, то этот критерий отсутствия связи тот же, что и критерий для двух пропорций. Рассчитываем частоту, которую мы ожидаем в каждой клетке таблицы сопряжённости, исходя из условия, что нулевая гипотеза верна. Как объяснено в главе 24, ожидаемая частота в отдельной клетке — произведение суммы частот в ряду и суммы частот в колонке, делённых на общее целое (объём всех наблюдений). Мы вычисляем статистику критерия, которая концентрируется на расхождении между наблюдаемыми и ожидаемыми частотами в каждой клетке таблицы. Если общее расхождение большое, то маловероятно, что нулевая гипотеза верна.

Таблица 25-1. Наблюдаемые (фактические) частоты в таблице сопряжённости $r \times c$

Градации строк	Градации колонок					
	Колонка 1	Колонка 2	Колонка 3		Колонка c	Всего
Строка 1	f_{11}	f_{12}	f_{13}		f_{1C}	R_1
Строка 2	f_{21}	f_{22}	f_{23}		f_{2C}	R_2
Строка 3	f_{31}	f_{32}	f_{33}		f_{3C}	R_3
Строка r	f_{r1}	f_{r2}	f_{r3}		f_{rC}	R_r
Всего	C_1	C_2	C_3	...	C_C	n

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: нет взаимосвязи в популяции между уровнями одного фактора и уровнями другого.
  - H_1: два фактора в популяции взаимосвязаны.
- 2. Отобрать необходимые данные из выборки пациентов.
- 3. Рассчитать величину статистики критерия, отвечающей H_0:

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

где O и E — наблюдаемая и ожидаемая частоты в каждой клетке таблицы. Статистика критерия подчиняется Хи-квадрат распределению Пирсона со степенями свободы $df=(r-1)\times(c-1)$.
  - Поскольку аппроксимация распределения Хи-квадрата оправдана, то, если число степеней свободы df больше единицы, у нас нет необходимости включать поправку Йэйтса на непрерывность (как это мы делали в главе 24).
- 4. Сравнить вычисленное значение статистики Хи-квадрат с табличным значением распределения вероятности. Обратитесь к табл. А3 в приложении А.
- 5. Интерпретировать величину достигнутого уровня значимости P и результат (наличие/отсутствие взаимосвязи)[1].

Если допущения не выполняются

Если более 20% ожидаемых частот меньше 5, мы пытаемся объединить два или более рядов и/или две или более колонок таблицы сопряжённости подходящим образом (например, так, чтобы это объединение имело научный смысл). Затем мы пересчитываем ожидаемые частоты этой уменьшенной таблицы и продолжаем уменьшение таблицы, если необходимо гарантировать, что условие $E \geq 5$ выполняется. Если мы уменьшили таблицу до таблицы 2×2, так что дальше её уменьшить нельзя, а всё ещё имеем маленькие ожидаемые частоты ($E < 5$), то применяем точный критерий Фишера для оценки точной величины уровня значимости P. Некоторые компьютерные программы рассчитывают точные значения критерия Фишера и для больших таблиц сопряжённости[2].

[1] Интерпретация наличия связи требует детального анализа её структуры в таблице частот, что сделать не столь просто, как может показаться, особенно в тех случаях, когда число градаций у обеих переменных более 3−1. В таком случае может помочь редукция таблицы, что можно сделать с помощью профессионального биостатистика. Прим. переводчика.

[2] По адресу http://www.biometrica.tomsk.ru/programm_stat.htm читатели могут познакомиться с такой программой и скачать её для работы. Прим. переводчика.

КРИТЕРИЙ ХИ-КВАДРАТ ПИРСОНА ДЛЯ УПОРЯДОЧЕННЫХ ГРАДАЦИЙ

Проблема

Иногда мы исследуем соотношение качественных данных, когда один из двух факторов имеет только две категории (например, наличие или отсутствие характеристики), а второй фактор может быть распределён в k альтернативных, взаимоисключающих градациях, которые в некотором смысле упорядочены. Например, одним фактором может быть реакция пациента на лечение (реагирует или нет), а упорядоченные категории другого фактора могут представлять собой 4 различных возрастных (в годах) группы 65–69, 70–74, 75–79 и 80 и более лет. Затем мы можем оценить, имеется ли в пропорциях тренд возрастания реакции пациентов при росте категорий второго фактора. Например, мы можем узнать, имеют ли тенденцию пропорции свойства, являющегося реакцией на лечение, увеличиваться, скажем, с возрастом.

Таблица 25-2. Наблюдаемые и ожидаемые (расчётные) частоты в таблице сопряжённости 2×k

Свойство	Столбец 1	Столбец 2	Столбец 3		Столбец k	Итого
Имеется	f_{11}	f_{12}	f_{13}	...	f_{1k}	R_1
Отсутствует	f_{21}	f_{22}	f_{23}		f_{2k}	R_2
Итого	C_1	C_2	C_3		C_k	n
Метки групп	w_1	w_2	w_3	...	w_k	

- **1. Определить нулевую и альтернативную гипотезы.**
  - ✧ H_0: нет тренда в пропорциях с характеристикой в популяции.
  - ✧ H_1: есть тренд пропорций в популяции.
- **2. Отобрать необходимые данные из выборки пациентов**. Оцениваются пропорции с характеристикой в каждой из k категорий. Назначают метки каждому из столбцов (категорий). Обычно это последовательные величины 1, 2, 3, .., k, но в зависимости от того, как мы классифицируем фактор колонки, они могут быть числами, которые некоторым образом предполагают относительные величины упорядоченных категорий (например, среднюю точку возрастного диапазона, определяющего каждую категорию) или тренд, который мы хотим изучить (например, линейный или квадратичный). Применение любых равномерно расположенных чисел (например, 1, 2, 3, .., k) позволяет изучить линейный тренд.
- **3. Вычислить значение статистики критерия, отвечающей H_0:**

$$\chi^2 = \frac{\left(\sum w_i f_{1i} - R_1 \sum \frac{w_i C_i}{n}\right)^2}{\frac{R_1}{n}\left(1-\frac{R_1}{n}\right)\left(\sum C_i w_i^2 - n\left(\sum \frac{w_i C_i}{n}\right)^2\right)}.$$

  - ✧ Используя условные обозначения из табл. 25-2, где суммы распространяются на все категории k. Статистика критерия подчиняется распределению Хи-квадрата со степенями свободы (k–1).
- **4. Сравнить величину статистики критерия с величинами из известного распределения вероятности.** Обратитесь к распределению Пирсона χ^2 в приложении А, табл. А3.
- 5. **Интерпретировать величину P и результаты.** Интерпретируйте величину P и рассчитайте доверительный интервал для каждой из k пропорций.

ПРИМЕР

Перекрёстное обследование было проведено среди популяции пожилых жителей Саутгемптона с целью измерения частоты сердечно-сосудистых заболеваний. Было опрошено в целом 259 лиц в возрасте от 65 до 95 лет. Во время опроса пациенты были распределены в 4 возрастные группы: (65–69, 70–74, 75–79, 80 и более лет). Для определения, различалась ли распространённость боли в груди в этих возрастных группах, мы использовали критерий Хи-квадрат Пирсона.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: нет связи между возрастом и болью в груди в популяции.
  - ✧ H_1: есть связь между возрастом и болью в груди в популяции.
- 2. Наблюдаемые частоты в процентах и ожидаемые частоты показаны в табл. 25-3.
- 3. Статистика критерия Хи-квадрат равна:

$$\chi^2 = \left[\frac{(15-9{,}7)^2}{9{,}7} + ... + \frac{(41-39{,}1)^2}{39{,}1}\right] = 4{,}839.$$

- 4. Обратившись к распределению Хи-квадрат в табл. А3 (приложение А) с 3 степенями свободы, получим P >0,10 (компьютерный расчёт даёт значение P =0,18).
- 5. Недостаточно данных для того, чтобы отбросить нулевую гипотезу относительно отсутствия связи между болью в груди и возрастом в популяции пожилых людей. Оценённые пропорции (95% доверительные интервалы) с болью в груди для этих четырёх возрастных групп начиная с самой молодой: 0,2 (0,11; 0,29); 0,12 (0,04; 0,19); 0,10 (0,02; 0,17) и 0,09 (0,02; 0,21).

Таблица 25-3. Наблюдаемые и ожидаемые (расчётные) частоты

Наличие боли в груди		Возраст, годы				
		65–69	**70–74**	**75–79**	**80 и более**	**Итого**
Есть	Наблюдаемая частота	15 (20,3%)	9 (11,5%)	6 (9,7%)	4 (8,9%)	34
	Ожидаемая частота	9,7	10,2	8,1	5,9	
Нет	Наблюдаемая частота	59 (79,7%)	69 (88,5%)	56 (90,3%)	41 (91,1%)	225
	Ожидаемая частота	64,3	67,8	53,9	39,1	
Итого		74	78	62	45	259

Группы в этом исследовании упорядочены, поэтому можно также проанализировать эти данные, используя критерий Хи-квадрат Пирсона для тренда, который принимает во внимание упорядоченность групп. Мы можем получить важный вывод от использования этого критерия, даже если общий критерий связи дал отрицательный результат. Припишем метки 1, 2, 3 и 4 соответственно каждой из четырёх возрастных групп и с учётом равной длины каждого интервала можем проверить на наличие линейного тренда.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - H_0: нет линейной связи между возрастом и болью в груди в популяции.
  - H_1: есть линейная связь между возрастом и болью в груди в популяции.
- 2. Данные отображены в вышеприведённой таблице. Приписываем метки 1, 2, 3, и 4 четырём возрастным группам соответственно.
- 3. Статистика критерия Пирсона Хи-квадрат равна:
- 4. Обратившись к распределению Хи-квадрат в приложении А, табл. А3 с 1 степенью свободы, получим $0{,}05<P<0{,}10$ (компьютерный расчёт даёт $P=0{,}052$).
- 5. Недостаточно данных, чтобы отбросить нулевую гипотезу об отсутствии линейной связи между болью в груди и возрастом в популяции пожилых людей. Однако достигнутый уровень значимости P очень близок к $P=0{,}05$, поэтому можно предположить, что доля пожилых людей с болью в грудной клетке уменьшается с увеличением возраста.

$$\chi^2=\frac{\left\{\left[(1\times 15)+\ldots+(4\times 4)\right]-34\times\left[\frac{1\times 74}{259}+\ldots+\frac{4\times 45}{259}\right]\right\}^2}{\frac{34}{259}\times\left(1-\frac{34}{259}\right)\times\left\{\left[(74\times 1^2)+\ldots+(45\times 4^2)\right]-259\times\left[\frac{1\times 74}{259}+\ldots+\frac{4\times 45}{259}\right]^2\right\}}=3{,}79$$

Данные взяты из Dewhurst G., Wood D.A., Walker E., et al. A population survey of cardiovascular disease in elderly people: design, methods and prevalence results. Age and Ageing, 1991. — Vol. 20. — P. 353–360.

26 Корреляция

ВВЕДЕНИЕ

Корреляционный анализ занимается измерением степени связи между двумя переменными x и y. Вначале мы предполагаем, что как x, так и y — количественные величины, например, рост и вес.

Предположим, что есть пара величин (x, y), измеренных у каждого из n пациентов в выборке. Мы можем отметить точку, соответствующую паре величин каждого пациента, на двухмерном графике рассеяния точек. Обычно мы располагаем переменную x на горизонтальной оси, а y — на вертикальной в той же диаграмме. Размещая точки для всех n пациентов, мы получаем график рассеяния точек, которые говорят о соотношении между этими двумя переменными.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ПИРСОНА

Говорят, что соотношение между x и y линейное, если прямая линия, проведённая через центральную часть скопления точек, даёт наиболее подходящую аппроксимацию наблюдаемого соотношения. Мы измеряем, как близко находятся наблюдения к прямой линии, которая лучше всего описывает их линейное соотношение путём расчёта произведения момента коэффициента корреляции Пирсона, обычно просто называемого коэффициент корреляции. Его истинная величина в популяции (генеральный коэффициент корреляции) ρ (греческая буква «ро») оценивается в выборке как r, которую обычно получают в результатах компьютерного расчёта:

$$r = \frac{\sum(x-\bar{x})\times(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}.$$

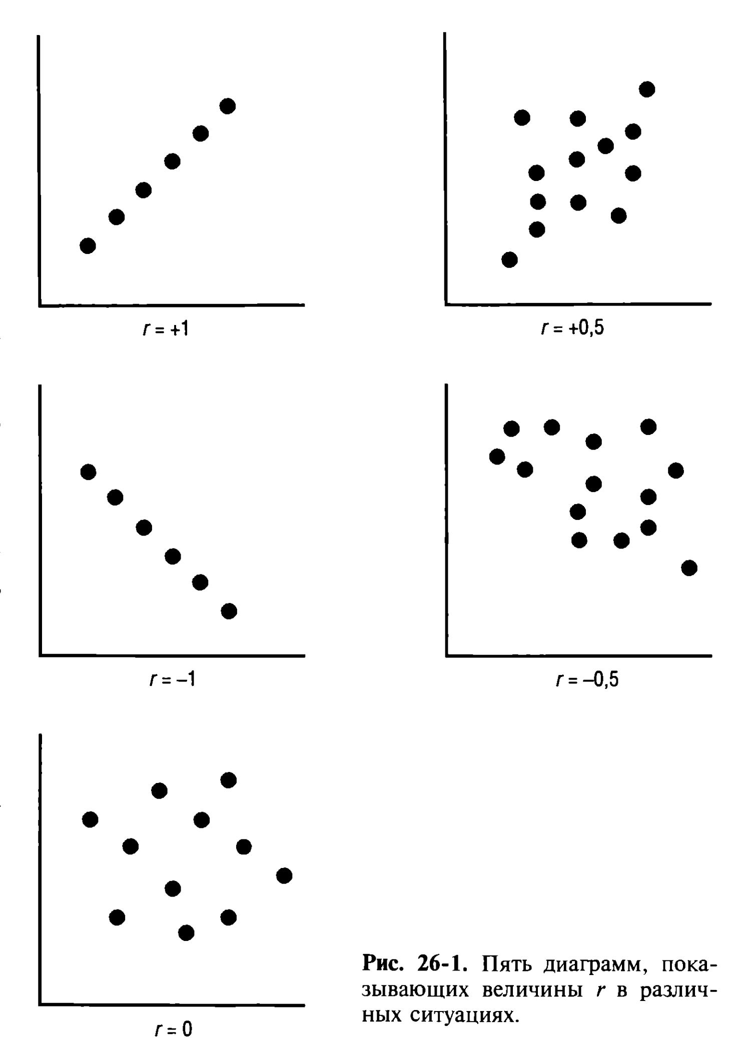


Рис. 26-1. Пять диаграмм, показывающих величины r в различных ситуациях.

Свойства

- r находится в интервале от -1 до $+1$.
- Его знак показывает, увеличивается ли одна переменная, по мере того как увеличивается другая (положительный r) или уменьшается (отрицательный r) (рис. 26-1).
- Его величина указывает, как близко расположены точки к прямой линии. В частности, если $r = +1$ или $r = -1$, то имеется абсолютная (функциональная) корреляция по всем точкам, лежащим на линии (практически это маловероятно); если $r=0$, то линейной корреляции нет (хотя может быть нелинейное соотношение). Чем ближе r к крайним точкам (±1), тем больше степень линейной связи (см. рис. 26-1).
- Коэффициент корреляции безразмерен, т.е. не имеет единиц измерения.
- Его величина действительна только в диапазоне значений x и y в выборке. Вы не можете заключить, что он будет иметь ту же величину при рассмотрении значений x или y, значительно больших, чем в выборке.
- x и y могут заменять друг друга, не влияя на величину r ($r_{xy} \equiv r_{yx}$).
- Корреляция между x и y не обязательно означает соотношение «причины и следствия».
- r^2 представляет собой долю вариабельности y, которая обусловлена линейным соотношением с x.

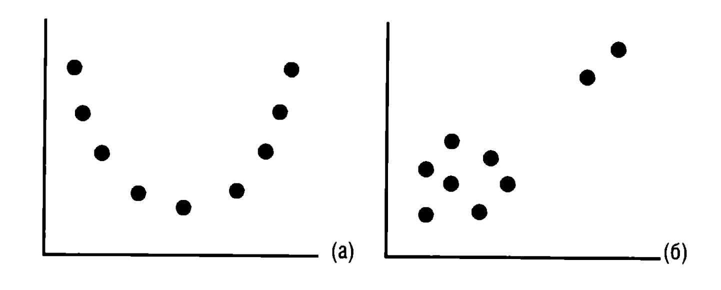
Когда не следует рассчитывать r

Расчёт r может ввести в заблуждение, если:

- получено нелинейное соотношение между этими двумя переменными (рис. 26-2, а), например, квадратичное соотношение;
- данные включают более одного наблюдения по каждому пациенту;
- присутствуют аномальные значения (рис. 26-2, б);
- данные содержат подгруппы пациентов, для которых средние уровни наблюдений, по крайней мере по одной из переменных, отличаются (рис. 26-2, в).

Критерий проверки гипотез для коэффициента корреляции Пирсона

Мы хотим знать, имеется ли какое-либо линейное соотношение между двумя количественными переменными. Выборка состоит из n независимых пар величин x и y. Мы допускаем, что по крайней мере одна из этих двух переменных нормально распределена.



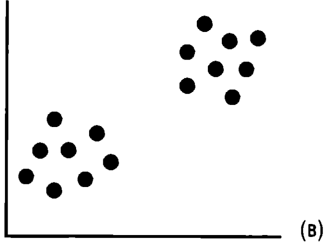


Рис. 26-2. Диаграммы, показывающие, когда не следует рассчитывать коэффициент корреляции (Пирсона). (а) Соотношение нелинейно, $r=0$. (б) При наличии выброса (выбросов). (в) Данные состоят из подгрупп.

- 1. Определить нулевую и альтернативную изучаемые гипотезы:
  - H_0: $\rho=0$;
  - H_1: $\rho\neq 0$.
- 2. Отобрать необходимые данные из выборки пациентов.
- 3. Вычислить значение статистики критерия, отвечающей H_0.
  - Вычислить r.
    - Если $n \leq 150$, r есть статистика критерия.
    - Если $n > 150$, вычислить t-статистику

$$T = r\sqrt{\frac{(n-2)}{(1-r^2)}},$$

имеющую t-распределение со степенями свободы $f=(n-2)$.
- 4. Сравнить величину статистики критерия с величинами из известного распределения вероятности:
  - Если $n \leq 150$, обратиться к приложению А, табл. А10.
  - Если $n > 150$, обратиться к приложению А, табл. А2.
- 5. Интерпретировать величину P и результаты.

Рассчитайте доверительный интервал для ρ. При условии, что обе переменные приблизительно нормально распределены, приблизительный 95% доверительный интервал для ρ:

$$\left(\frac{e^{2z_1}-1}{e^{2z_1}+1};\ \frac{e^{2z_2}-1}{e^{2z_2}+1}\right),$$

где

$$z_1 = z - \frac{1{,}96}{\sqrt{n-3}},\quad z_2 = z + \frac{1{,}96}{\sqrt{n-3}}$$

и $$z = 0{,}5\ln\left[\frac{(1+r)}{(1-r)}\right].$$

Отметьте, что, если объём выборки большой, H_0 можно отбросить, даже если r довольно близко к нулю. Альтернативно, даже если r большое, H_0 можно не отбрасывать, если размер выборки маленький. По этой причине особенно полезно рассчитать r^2, долю общей вариации, объяснённую этим соотношением. Например, если $r=0{,}40$, тогда $P<0{,}05$ для размера выборки 25, но это соотношение объясняет только 16% ($=0{,}40^2\times 100$) вариабельности одной из переменных.

РАНГОВЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ СПИРМЭНА[1]

Мы вычисляем ранговый коэффициент корреляции Спирмэна, непараметрический аналог корреляционного коэффициента Пирсона, если верны одно или два из следующих условий:

- по крайней мере одна из переменных x или y измеряется в ранговой (порядковой) шкале;
- ни x, ни y не распределены нормально;
- размер выборки небольшой;
- требуется измерение связи между двумя переменными, когда их соотношение нелинейно.

Вычисление

Оценка популяционной величины рангового коэффициента корреляции Спирмэна ρ_s по выборочному значению r_s.

- 1. Расположите величины x в возрастающем порядке, начиная с наименьшей величины, и придайте им последовательные ранги (номера 1, 2, 3, .., n). Равные величины получают среднее значение из рангов, какое получили бы эти значения, если бы не было такого ограничения.
- 2. Подобным образом придайте ранги величинам y.
- 3. r_s — коэффициент корреляции Пирсона между рангами x и y.

Свойства и критерий проверки гипотез

Они такие же, как и для коэффициента корреляции Пирсона (r заменяют на r_s). Однако:

- r_s даёт измерение связи (необязательно линейной) между x и y;
- тестируя нулевую гипотезу о том, что $\rho_s=0$, обратитесь к табл. А11 в приложении А, если размер выборки менее или равен 10;
- мы не вычисляем значение r_s^2 (оно не представляет собой долю общего вариации одной переменной, которую можно объяснить изменением другой переменной).

[1] В англоязычной литературе часто используют вместо буквы ρ название этой буквы в латинской графике — Rho, соответственно записывая в тексте «Spearman's Rho». Прим. переводчика.

ПРИМЕР

Как часть исследования по изучению факторов, связанных с изменениями кровяного давления у детей, была собрана информация о демографических факторах и образу жизни, а также клинические и антропометрические измерения у 4245 детей в возрасте 5–7 лет. Соотношение между ростом (см) и САД (мм рт.ст.) в выборке из 100 детей показано на двухмерном графике (рис. 28-1); наблюдаем тенденцию: у более высоких детей в выборке более высокое артериальное давление. Был изучен корреляционный коэффициент Пирсона между этими двумя переменными. Приложение В содержит результаты компьютерного расчёта из анализа.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.

  H_0: величина корреляционного коэффициента Пирсона ρ в популяции равна нулю.

  H_1: величина корреляционного коэффициента Пирсона ρ в популяции не равна нулю.
- 2. Мы можем показать (рис. 37-1), что выборочные значения как веса, так и САД приблизительно нормально распределены.
- 3. Мы рассчитываем r как 0,33. Проверим статистическую значимость, поскольку $n \leq 150$. Обратимся к приложению А, табл. А10 с размером выборки в 100: $P < 0,001$.
- 4. Есть убедительные аргументы, чтобы отбросить нулевую гипотезу; мы заключаем, что между САД и ростом в популяции таких детей имеется линейное соотношение. Однако $r^2=0,33\times0,33=0,11$, поэтому, несмотря на статистически значимый результат, соотношение между весом и САД объясняет только небольшой процент, 11%, изменения САД.
- 5. Чтобы определить 95% доверительный интервал для генерального коэффициента корреляции, мы вычисляем:

$$z=0,5\ \ln(1,33/0,67)=0,3428;$$

$$z_1=0,3428-1,96/9,849=0,1438;$$

$$z_2=0,3428+1,96/9,849=0,5418.$$

Таким образом, доверительный интервал варьирует от

$$\frac{\left(e^{2\times0,1438}-1\right)}{\left(e^{2\times0,1438}+1\right)} \text{ до } \frac{\left(e^{2\times0,5418}-1\right)}{\left(e^{2\times0,5418}+1\right)},$$

$$\text{т.е. от } \frac{0,33}{2,33} \text{ до } \frac{1,96}{3,96}.$$

Мы на 95% уверены, что ρ лежит в интервале между 0,14 и 0,49.

Итак, при условии, что каждая переменная нормально распределена, ранговый корреляционный коэффициент Спирмэна между этими переменными даёт достаточно близкую оценку, равную 0,32. Для проверки H_0: $\rho_s=0$ мы найдём эту величину в приложении А, табл. А10 и определим $P < 0,001$.

Данные любезно предоставлены Ms O. Papacosta and Dr p. Whincup, Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK.

27 Теория линейной регрессии

ЧТО ТАКОЕ ЛИНЕЙНАЯ РЕГРЕССИЯ?

Для изучения соотношения между двумя непрерывными переменными *x* и *y* измеряем величины *x* и *y* для каждого из *n* пациентов в выборке. Размещаем точки на двухмерном графике рассеяния и говорим, что мы имеем линейное соотношение, если данные аппроксимируются прямой линией. Если мы полагаем, что *y* зависит от *x*, причём изменения в *y* вызываются именно изменениями в *x*, мы можем определить линию регрессии (регрессия *y* на *x*), которая лучше всего описывает прямолинейное соотношение между этими двумя переменными. Вообще, регрессию описываем как одномерный анализ, потому что нас интересует в анализе только одна переменная *x*. Этот анализ противопоставляется множественной регрессии, в которой участвуют две или более объясняющих переменных.

ЛИНИЯ РЕГРЕССИИ

Математическое уравнение, которое оценивает линию простой линейной регрессии:

$$Y=a+bx.$$

x называется независимой, предиктором, или объясняющей переменной.

Для данной величины *x*, *Y* — значение переменной *y* (называемой зависимой, выходной переменной, или переменной отклика), которое расположено на линии оценки. Это есть значение, которое мы ожидаем для *y* (в среднем), если мы знаем величину *x*, и называется она «предсказанное значение *y*».

- *a* — свободный член (пересечение) линии оценки; это значение *Y*, когда *x*=0 (рис. 27-1).
- *b* — угловой коэффициент или градиент оценённой линии; он представляет собой величину, на которую *Y* увеличивается в среднем, если мы увеличиваем *x* на одну единицу (см. рис. 27-1).
- *a* и *b* называют коэффициентами регрессии оценённой линии, хотя этот термин часто используют только для *b*. Мы покажем, как оценить эти коэффициенты, в главе 28. Простую линейную регрессию можно расширить, включив в нее более одной независимой переменной; в этом случае она известна как множественная линейная регрессия.

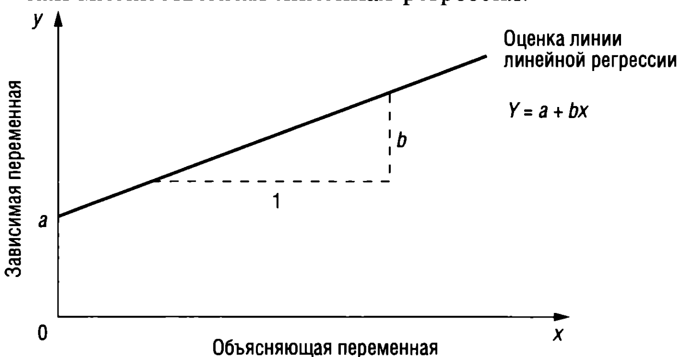


Рис. 27-1. Линия линейной регрессии, показывающая пересечение *a* и угловой коэффициент *b* (величину возрастания *Y* при увеличении *x* на одну единицу).

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Мы выполняем регрессионный анализ, используя выборку наблюдений, где *a* и *b* — выборочные оценки истинных (генеральных) параметров, α и β, которые определяют линию линейной регрессии в популяции (генеральной совокупности). *a* и *b* определяют методом наименьших квадратов (чаще называемым обычным МНК) таким образом, что «подгонка» линии *Y*=*a*+*bx* к точкам на двухмерном графике рассеяния оптимальна. Мы оцениваем это, рассматривая остатки (вертикальное расстояние каждой точки от линии, например, остаток = наблюдаемому *y* — предсказанный *Y*, рис. 27-2). Линию лучшей подгонки выбирают так, чтобы сумма квадратов остатков была минимальной.

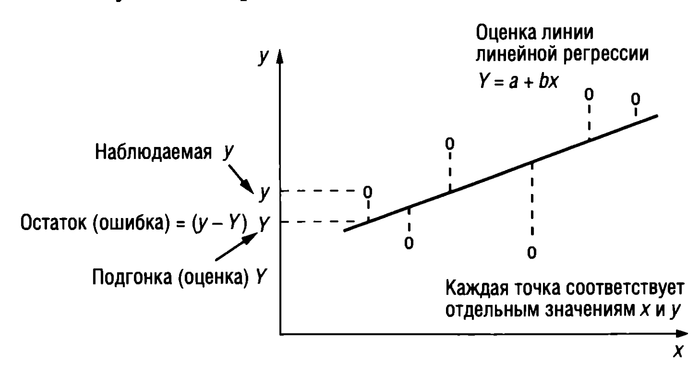


Рис. 27-2. Линия линейной регрессии с изображёнными остатками (вертикальные пунктирными линии) для каждой точки.

ДОПУЩЕНИЯ

- 1. Между *x* и *y* существует линейное соотношение.
- 2. Наблюдения в выборке независимы. Наблюдения независимы, если по каждому пациенту есть не более одной пары наблюдений.
- 3. Для каждой величины *x* имеется распределение величин *y* в популяции; это распределение нормально[1]. Среднее этого распределения величин *y* лежит на истинной линии регрессии (рис. 27-3).

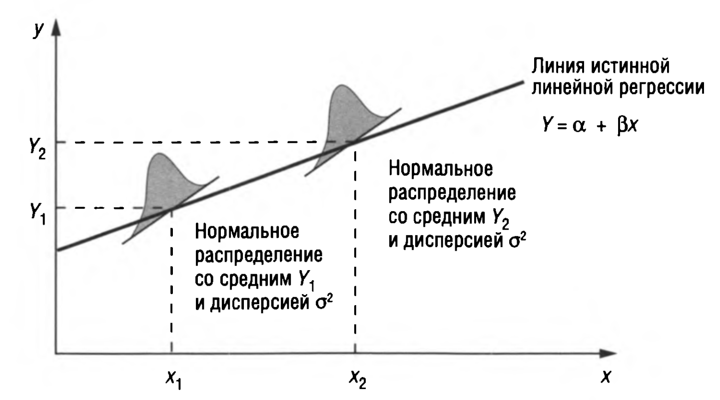


Рис. 27-3. Иллюстрация предположений, сделанных в линейной регрессии.

[1] Данное требование необязательно. Гораздо важнее выполнение нормальности распределения ошибок — разностей между фактическим и предсказанным значением зависимой переменной *y*. Прим. переводчика.

- 4. Вариабельность распределения величин у в популяции одинакова для всех величин x, т.е. дисперсия σ^2 постоянна (см. рис. 27-3).
- 5. Переменную x можно измерить без ошибки. Отметьте, что мы не делаем никаких допущений о распределении переменной x.

Многие допущения, которые лежат в основе регрессионного анализа, относятся к распределению y в популяции для заданного значения x, но они могут быть сформулированы в рамках терминов остатков. Легче проверить допущения изучая остатки, чем величины y.

ТАБЛИЦА ДИСПЕРСИОННОГО АНАЛИЗА

Описание

Обычно компьютерный результат в регрессионном анализе содержит таблицу дисперсионного анализа. В дисперсионном анализе общая вариация интересующих нас переменных, в данном случае y, разделяется на отдельные компоненты. Из-за линейного соотношения y и x мы ожидаем, что y изменяется, по мере того как изменяется x, и называем это вариацией, которая обусловлена или объясняется регрессией. Остаточная вариация называется остаточной ошибкой, или необъяснённой вариацией. Остаточная вариация должна быть как можно меньше. Если это так, то бóльшая часть вариации y будет объясняться регрессией, а точки будут лежать близко к линии регрессии, т.е. линия хорошо соответствует данным.

Цели

Таблица дисперсионного анализа позволяет сделать следующие выводы.

1. Оценить, насколько хорошо линия регрессии соответствует точкам данных. Из информации, данной в таблице, можно рассчитать долю общей дисперсии y, которая объясняется регрессией. Эта доля, которую называют коэффициентом детерминации, обычно выражают через процентное соотношение и обозначают R^2 (в парной линейной регрессии это величина r^2, квадрат коэффициента корреляции), позволяет субъективно оценить качество согласия уравнения регрессии.

2. Проверить нулевую гипотезу о том, что генеральный угловой коэффициент линии регрессии β равен нулю; указать на статистическую значимость результата о существовании линейной регрессии между x и y.

3. Получить оценку дисперсии остатков. Нам нужна эта оценка для проверки гипотез об угловом коэффициенте или свободного члена и при расчёте доверительных интервалов для этих параметров и предсказанных величин y.

Детали более общих процедур мы даём в главе 28.

РЕГРЕССИЯ К СРЕДНЕМУ

Статистическое использование слова «регрессия» исходит из явления, известного как регрессия к среднему, приписываемого сэру Френсису Гальтону (1889). Он показал, что, хотя высокие отцы имеют тенденцию иметь высоких сыновей, средний рост сыновей меньше, чем у их высоких отцов. Средний рост сыновей «регрессировал» или «двигался вспять» к среднему росту всех отцов в популяции. Таким образом, в среднем высокие отцы имеют более низких (но всё-таки высоких) сыновей, а низкие отцы имеют сыновей более высоких (но всё-таки довольно низких).

Мы наблюдаем регрессию к среднему при скрининге и клинических исследованиях, когда подгруппа пациентов может быть выбрана для лечения потому, что их уровни определённой переменной, скажем, холестерина, крайне высоки (или низки). Если это измерение через некоторое время повторяется, средняя величина второго считывания для подгруппы обычно меньше, чем при первом считывании, имея тенденцию (т.е. регрессируя) к среднему, подобранному по возрасту и полу в популяции, независимо от лечения, которое они могут получить. Пациенты, набранные в клиническое исследование на основе высокого уровня холестерина при их первом осмотре, таким образом, вероятно, покажут в среднем падение уровня холестерина при втором осмотре, даже если в этот период они не лечились.

28 Проведение анализа линейной регрессии

ЛИНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

После отбора выборки объёмом n из популяции и вычерчивания двухмерного графика для подтверждения, что данные аппроксимируются прямой линией, мы оцениваем регрессию y на x как:

$$Y = a + bx,$$

- Y — предсказанная величина y;
- a — пересечение;
- b — угловой коэффициент, который представляет собой среднее изменение для Y при единичном изменении x.

НАНЕСЕНИЕ ЛИНИИ

Чтобы нанести линию $Y=a+bx$ на двухмерном графике, мы выбираем три значения переменной x: x_1, x_2 и x_3 внутри интервала изменения. Подставляем x_1 в уравнение для получения соответствующего значения Y, а именно $Y_1=a+bx_1$; это предсказанная величина по значению x_1, которая соответствует наблюдаемой величине y_1. Повторяем процедуру для x_2 и x_3 для получения соответствующих величин Y_2 и Y_3. Располагаем эти точки на двухмерном графике и соединяем их для создания прямой линии.

ПРОВЕРКА ДОПУЩЕНИЙ

Для каждой наблюдаемой величины x остаток равен разнице y и соответствующего предсказанного Y. Каждый остаток может быть положительным или отрицательным. Можно использовать остатки для проверки следующих предположений, лежащих в основе линейной регрессии.

- 1. Между x и y существует линейное соотношение: для любых пар (x; y) данные должны аппроксимировать прямую линию, или надо нанести вместо них остатки (мы должны наблюдать случайный разброс точек, а не какую-либо систематическую картину).
- 2. **Наблюдения независимы**, если имеется не более одной пары наблюдений по каждому пациенту.
- 3. Остатки нормально распределены с нулевым средним значением: начертите гистограмму, график «стебель и листья», график «ящик с усами» или нормальный график остатков и посмотрите результат.
- 4. Остатки имеют одну и ту же вариабельность (постоянную дисперсию) для всех предсказанных величин y: нанесите остатки против предсказанных величин Y от y; мы должны наблюдать случайное рассеяние точек. Если график рассеяния остатков увеличивается или уменьшается с увеличением Y, то это допущение не выполняется.
- 5. Переменная x может быть измерена без ошибки.

Невозможность удовлетворить допущения

Если допущения линейности, нормальности и/или постоянной дисперсии сомнительны, мы можем преобразовать x или y и рассчитать новую линию регрессии, для которой эти допущения удовлетворяются. Не всегда можно найти удовлетворительное преобразование. Допущения линейности и независимости наиболее важны. Если вы сомневаетесь относительно допущений нормальности и/или постоянной дисперсии, вы можете продолжать, но величины P в критериях проверки гипотез и оценки стандартных ошибок могут быть искажены. Заметьте, что переменная x редко измеряется без какой-либо ошибки; при условии, что ошибка небольшая, это обычно приемлемо, поскольку её влияние на результаты минимально.

АНОМАЛЬНЫЕ ЗНАЧЕНИЯ (ВЫБРОСЫ) И ТОЧКИ ВЛИЯНИЯ

Влиятельное наблюдение, если оно опущено, изменяет одну или больше оценок параметров модели (т.е. угловой коэффициент или свободный член). Формальные методы обнаружения обсуждены кратко в главе 29. Если эти методы недоступны, вероятно, придётся полагаться на интуицию.

Выброс (наблюдение, которое противоречит большинству значений в наборе данных) может быть/не быть влительным наблюдением и может хорошо обнаруживаться визуально, при осмотре двухмерной диаграммы рассеяния или графика остатков. И для выбросов, и для влиятельных наблюдений (точек) используют модели как с подозрительными наблюдениями, так и без них и обращают внимание на изменение оценок (коэффициентов регрессии). Не отбрасывайте выбросы или точки влияния рутинно, поскольку простое их игнорирование может повлиять на ваши заключения. Всегда изучайте причины появления этих выбросов и сообщайте о них.

ОЦЕНКА КАЧЕСТВА СОГЛАСИЯ

Насколько хорошо линия регрессии согласуется с данными, можно судить, рассчитав R^2 (обычно выраженный в процентах и называемый коэффициентом детерминации), который равняется квадрату коэффициента корреляции. Он представляет собой долю или процент дисперсии y, который можно объяснить связью с x. Разность $(100-R^2)$ представляет собой процент дисперсии y, который нельзя объяснить этим взаимодействием. Нет формального теста для оценки R^2; мы вынуждены положиться на субъективное суждение, чтобы определить согласие линии регрессии[1].

ИССЛЕДОВАНИЕ УГЛОВОГО КОЭФФИЦИЕНТА

Если угловой коэффициент линии равен нулю, между x и y нет линейного соотношения: изменение x не влияет на y. Имеется два подхода с идентичными резуль-

[1] Для проверки статистической значимости коэффициента детерминации R^2 используют F-критерий Фишера. Процедура проверки и соответствующие формулы приведены в многочисленных учебниках и пособиях (см. http://www.biometrica.tomsk.ru/razdel_2_1.htm – http://www.biometrica.tomsk.ru/razdel_2_12.htm). Прим. переводчика.

татами для тестирования нулевой гипотезы о том, что истинный угловой коэффициент β равен нулю.

- Проверьте отношение F (равное отношению «необъяснённых» к «объяснённым» средним квадратам отклонений) в таблице дисперсионного анализа. Оно подчиняется F-распределению и имеет (1; n−2) степеней свободы соответственно в числителе и знаменателе.
- Вычислите статистику критерия, равного отношению b/SE(b), которая подчиняется t-распределению с (n−2) степенями свободы, где SE(b) — стандартная ошибка коэффициента b.

В любом случае результат статистически значимый, обычно если P <0,05, приводит к отбрасыванию нулевой гипотезы.

Мы рассчитываем 95% доверительный интервал для генерального углового коэффициента β как $b \pm t_{0,05}\mathrm{SE}(b)$, где $t_{0,05}$ — процентная точка t-распределения со степенями свободы (n−2), что даёт вероятность двустороннего критерия в 0,05. Это тот интервал, который содержит генеральный угловой коэффициент с вероятностью в 95%. Для больших выборок, скажем, $n \geq 100$, мы можем аппроксимировать $t_{0,05}$ значением 1,96.

Анализ регрессии редко выполняют вручную; компьютерный выход в большинстве статистических пакетов программ даст всю эту информацию.

ПРИМЕНЕНИЕ ЛИНИИ РЕГРЕССИИ ДЛЯ ПРОГНОЗА

Можно применять регрессионную линию для прогноза значения y по значению x в пределе наблюдаемого диапазона (никогда не экстраполируйте вне этих пределов). Мы предсказываем среднюю величину y для пациентов, которые имеют определённое значение x, путём подстановки этого значения x в уравнение линии регрессии. Итак, если $x=x_0$, прогнозируем у как $Y_0=a+bx_0$. Используем эту предсказанную величину и её стандартную ошибку, чтобы оценить доверительный интервал для истинной средней величины y в популяции. Повторение этой процедуры для различных величин x позволяет построить доверительные границы для этой линии. Это полоса или область, которая содержит истинную линию, например, с 95% доверительной вероятностью. Подобным образом можно рассчитать более широкую область, внутри которой, как мы ожидаем, лежит наибольшее число (обычно 95%) наблюдений.

Полезные формулы для ручных вычислений:

$$\bar{x} = \sum x/n \quad \text{и} \quad \bar{y} = \sum y/n,$$

$$a = \bar{y} - b\bar{x},$$

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2},$$

$$s_{\mathrm{res}}^2 = \frac{\sum (y-Y)^2}{(n-2)}, \text{ оценка дисперсии остатков.}$$

$$\mathrm{SE}(b) = \frac{s_{\mathrm{res}}}{\sqrt{\sum (x-\bar{x})^2}}.$$

ПРИМЕР

Соотношение между ростом (измеренным в см) и САД (измеренным в мм рт.ст.) у детей, описанное в главе 26, показано на рис. 28-1. Мы провели анализ парной линейной регрессии зависимости САД от роста. Допущения, лежащие в основе этого анализа, подтверждены на рис. 28-2—28-4. Типичный результат компьютерного анализа показан в приложении В. Имеется существенное линейное соотношение между ростом и САД, что можно видеть с помощью значимого F-отношения в таблице дисперсионного анализа в приложении В (F =12,03 со степенями свободы 1 и 98 соответственно в числителе и знаменателе, P=0,0008). R^2 модели равно 10,9%. Только около одной десятой дисперсии САД можно, таким образом, объяснить моделью, т.е. различиями в росте детей.

Компьютерный результат даёт следующую информацию.

Таблица 28-1. Соотношение между ростом и систолическим артериальным давлением у детей

Переменная	Параметры оценки (коэффициенты регрессии)	Стандартная ошибка параметра	Статистика критерия	Достигнутый уровень значимости (P-значение)
Свободный член	46,2817	16,7845	2,7574	0,007
Рост	0,4842	0,1396	3,4684	0,0008

Оценка параметра для «свободного члена» соответствует значению «a», а оценка параметра для «роста» соответствует значению «b» (угловой коэффициент линии регрессии). Итак, уравнение линии оценённой регрессии имеет следующий вид:

САД=46,28+0,48×рост.

В этом примере свободный член не представляет интереса (рост, равный нулю, явно вне диапазона величин, наблюдаемых в исследовании). Однако мы можем интерпретировать угловой коэффициент; предсказано, что у этих детей САД увеличивается в среднем на 0,48 мм рт.ст. при увеличении роста на один сантиметр. P=0,0008 для критерия проверки гипотезы по росту (H_0: генеральный угловой коэффициент равен нулю), как и ожидалось, идентичен уровню значимости, полученному в таблице дисперсионного анализа в приложении В.

Так как объём выборки большой (равен 100), мы можем аппроксимировать $t_{0,05}$ значением 1,96 и рассчитать 95% доверительный интервал для генерального углового коэффициента таким образом:

$$b \pm 1{,}96 \times SE(b) = 0{,}48 \pm (1{,}96 \times 0{,}14).$$

Итак, 95% доверительный интервал для углового коэффициента варьирует изменением от 0,21 до 0,75 мм рт.ст. на один сантиметр увеличения роста.

Этот доверительный интервал не включает ноль, подтверждая вывод о том, что угловой коэффициент статистически значимо отличается от нуля. Мы можем применить уравнение регрессии для предсказания САД, которое мы ожидаем у ребёнка при данном росте. Например, ребёнок ростом 115 см имеет предсказанное САД, равное 46,28+(0,48×115)=101,48 мм рт.ст; ребёнок ростом 130 имеет предсказанное САД, равное 46,28+(0,48×130)=108,68 мм рт.ст.

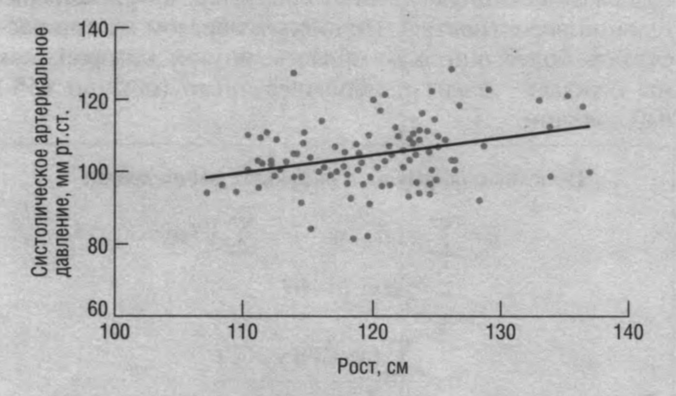


Рис. 28-1. Двухмерный график, показывающий соотношение между систолическим артериальным давлением и ростом. Изображена оценённая линия регрессии, систолическое артериальное давление =46,28+0,48×рост.

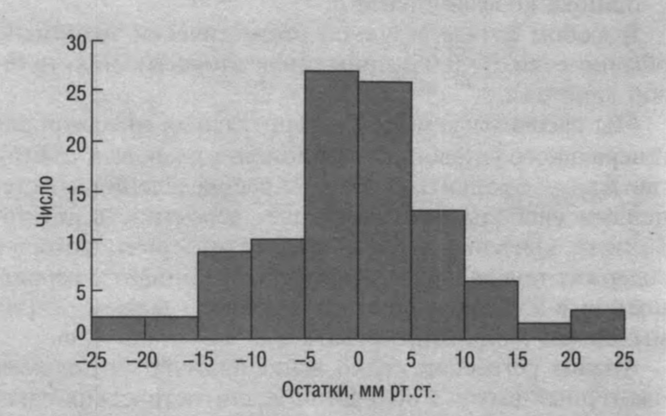


Рис. 28-3. Распределение остатков приблизительно нормально.

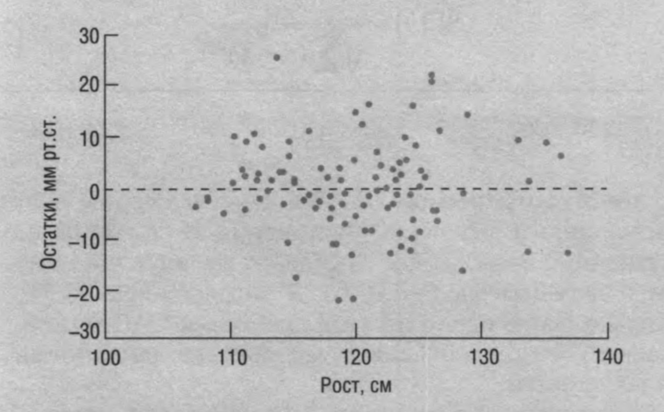


Рис. 28-2. На этой диаграмме нет явного соотношения, означающего, что есть линейное соотношение между ростом и остатками.



Рис. 28-4. Нет тенденции для остатков систематически увеличиваться или уменьшаться с предсказанными значениями. Следовательно, допущение постоянной дисперсии выполняется.

29 Множественная линейная регрессия

ЧТО ЭТО?

Можно интересоваться результатом влияния нескольких независимых переменных x_1, x_2, .., x_k на переменную отклика y. Если мы полагаем, что эти x могут быть взаимозависимы, то не должны смотреть по отдельности на эффект изменения значения одного x на y, но должны одновременно принимать во внимание величины всех других x. Например, поскольку между ростом и массой тела ребёнка существует сильная зависимость, можно поинтересоваться, изменяется ли также соотношение между ростом и САД, если принять во внимание также и массу тела ребёнка. Множественная линейная регрессия позволяет изучить совместный эффект этих нескольких независимых переменных на y. Отметьте, что хотя объясняющие переменные (предикторы) называются независимыми, это не вполне корректное название, поскольку они могут быть зависимыми.

Мы берём выборку из n пациентов и измеряем значение каждой из переменных у каждого пациента. Уравнение множественной линейной регрессии, которое оценивает эти соотношения в популяции, такое:

$$Y=a+b_1x_1+ b_2x_2+...+ b_kx_k,$$

- x_i — независимая переменная (предиктор), или ковариата (i=1, 2, 3, .., k);
- Y — предсказываемая или ожидаемая величина y, которая соответствует конкретному набору значений x_1, x_2, .., x_k;
- a — постоянная величина, иногда называемая пересечением (свободный член); значение переменной Y, когда все x равны нулю;
- b_1, b_2, .., b_k — оценки отдельных коэффициентов регрессии; b_1 представляет собой значение, на которое в среднем увеличивается Y, если мы увеличиваем x_1 на одну единицу, но сохраняем постоянными все другие x (т.е. согласовываем с ними). Если имеется соотношение между x_1 и другими предикторами, b_1 отличается от оценки коэффициента регрессии, полученного для парной регрессии y от x, поскольку последний вариант не учитывает другие предикторы. В этом случае b_1 представляет собой эффект x_1 на x, который не зависит от других предикторов.

Конечно, вы хотите провести анализ множественной регрессии на компьютере, поэтому мы опускаем формулы для получения оценок этих параметров.

ЗАЧЕМ ЭТО ДЕЛАТЬ?

Чтобы иметь возможность:

- определить интенсивность линейной взаимосвязи каждой независимой переменной с зависимой, с учётом влияния других переменных[1];
- предсказать величину зависимой переменной по значениям независимых.

ДОПУЩЕНИЯ

Допущения при множественной регрессии такие же (если мы заменяем «x» на «каждый из x»), что и при парной линейной регрессии, и проверяют их таким же образом. Особенно важно требование удовлетворить допущениям линейности или независимости. Мы можем преобразовать переменную y и/или несколько, или все предикторы x, если допущения вызывают сомнение, а затем повторить анализ (включая проверку допущений) на преобразованных данных.

КАЧЕСТВЕННЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

Можно провести анализ множественной регрессии с применением качественных независимых переменных. В частности, если мы имеем бинарную переменную x_1 (например, мужчина=0, женщина=1) и, увеличивая x_1 на одну единицу, «меняем» мужчин на женщин. Таким образом, b_1 представляет собой разницу в оценённых средних величинах у между женщинами и мужчинами, после подгонки для других x.

Если мы имеем номинальную переменную, которая имеет более двух категорий ответа (см. главу 1), мы должны создать ряд новых (искусственных) индикаторных переменных[2]. Некоторые компьютерные программы сделают это автоматически. Однако если есть ординальная переменная и её трём или более категориям можно присвоить величины на содержательной линейной шкале (например, социальные классы 1–5), то можно применить эти величины прямо в уравнении множественной регрессии.

Вообще, для номинальной переменной с k категориями мы создаём k–1 бинарных фиктивных переменных. Выбираем одну из категорий, чтобы представить её в качестве начальной категории (градации, уровня), и каждая фиктивная переменная позволяет нам сравнивать одну из оставшихся k–1 качественных пере-

---

[1] Важным моментом для множественной регрессии является ранжирование независимых предикторов по силе их влияния на зависимую переменную. Для этой цели используют так называемые безразмерные бэта-коэффициенты, ранжируя модули которых, можно проранжировать предикторы по степени влияния на зависимую переменную. Технология вычисления таких безразмерных коэффициентов несложна и описана во многих изданиях по регрессии, в том числе приведённых в электронной библиотеке по адресам http://www.biometrica.tomsk.ru/razdel_2_1.htm — http://www.biometrica.tomsk.ru/razdel_2_12.htm Прим. переводчика.

[2] Armitage P., Berry G. and Matthews J.N.S. Statistical Methods in Medical Research. — 4rd edn. — Blackwell (UK), 2001.

менных с начальной категорией. Например, мы можем интересоваться сравнением средних уровней систолического давления крови у пациентов, живущих в четырёх странах Европы (Нидерланды, Великобритания, Испания и Франция). Предположим, что мы выбираем начальную категорию Нидерландами. Производим одну бинарную переменную, чтобы идентифицировать проживание в Великобритании; эта переменная имеет значение 1, если пациент живёт в Великобритании, и 0 — в ином случае. Аналогичным образом мы сгенерируем бинарные переменные и для того, чтобы идентифицировать проживание в Испании и Франции. По умолчанию проживание в Нидерландах может быть тогда идентифицировано, когда пациент будет иметь значение 0 для каждой из трёх бинарных переменных. В множественном линейном регрессионном анализе регрессионный коэффициент для каждой из трёх других стран представляет величину, на которую Y (систолическое давление крови) проживающих в конкретной стране отличается в среднем от показателя пациентов, проживающих в Нидерландах. Свободный член (пересечение) уравнения регрессии предусматривает оценку среднего систолического давления крови проживающего в Нидерландах (когда все другие предикторы имеют значение равное нулю). Некоторые компьютерные пакеты создают фиктивные переменные автоматически, как только вы определили, что переменная категорическая.

Если мы имеем порядковый предиктор и его трём или более градациям могут быть указаны значения на поддающейся интерпретации линейной шкале (например, социальные классы 1–5), то можем или использовать эти значения непосредственно в множественном линейном уравнении регрессии, или сгенерировать ряд фиктивных переменных как для номинальной переменной (но это не производит упорядоченных категорий).

КОВАРИАЦИОННЫЙ АНАЛИЗ

Расширенный дисперсионный анализ (ANOVA) — это ковариационный анализ, в котором мы сравниваем интересующие нас отклики между группами пациентов (например, две или более групп лечения), когда принимаются во внимание другие переменные по каждому пациенту. Такие данные можно проанализировать, применяя способы множественной регрессии, путём создания одной или более бинарных переменных для разделения групп. Так, если мы хотим сравнить средние значения в двух группах лечения, контролируя эффект переменных x_2, x_3, .., x_k (т.е. рост, масса тела и т.д.), мы создаём бинарную переменную x_1, чтобы представить «лечение» (например, $x_1=0$ для лечения A, $x_1=1$ для лечения B). В уравнении множественной регрессии b_1 — оценка разницы между средними значениями y при лечении методами B и A, адаптированными для остальных x.

Анализ ковариаций — предпочтительный вид анализа для рандомизированных контролируемых исследований, сравнивающий виды лечения, когда по каждому пациенту есть данные на протяжении всего исследования. В этом случае переменная ответа, y — текущее измерение, а из двух объясняющих переменных (предикторов) в модели регрессии одна бинарная переменная x_1 представляет вид лечения, а переменная x_2 — значение основного признака пациента в начале исследования. Этот подход в целом лучше (т.е. имеет большую мощность, см. главу 36), чем использование изменения основного признака или изменения в процентах этого признака как зависимой переменной.

ВЫБОР НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ

Приближённое правило гласит, что мы не должны проводить анализ множественной регрессии, если количество переменных больше, чем количество пациентов, делённое на 10. Большинство компьютерных пакетов программ автоматически выбирают переменные, например, используя пошаговый отбор. Это особенно полезно, когда взаимосвязано много независимых переменных. Особая проблема возникает, когда имеется мультиколлинеарность, например, когда пары независимых переменных имеют крайне высокую корреляцию.

АНАЛИЗ

Большинство компьютерных результатов регрессионного анализа содержит следующие пункты.

- 1. Оценка качества согласия. Скорректированный R^2 представляет собой долю дисперсии y, которую можно объяснить его связью с предикторами. R^2 корректируется так, чтобы можно было сравнивать модели с различным количеством независимых переменных. Если он имеет низкое значение (субъективное мнение), модель имеет плохое согласие. Качество согласия особенно важно, когда используют уравнение множественной регрессии для прогноза.
- 2. F-критерий в таблице ANOVA. Этот критерий проверяет нулевую гипотезу о том, что все частные коэффициенты регрессии в популяции β_1, β_2, .., β_k равны нулю. Значимый результат (например, $P<0{,}05$) означает, что между y и по крайней мере одним из x имеется линейное соотношение.
- 3. t-критерий каждого отдельного коэффициента регрессии, β_i (i=1, 2, .., k). Каждый t-критерий относится к одной независимой переменной и важен, если мы хотим определить, влияет ли эта независимая переменная на переменную ответа, контролируя эффекты других ковариат. Для гипотезы H_0: $\beta_i=0$, мы вычисляем статистику критерия, равную $b_i/\mathrm{SE}(b_i)$, которая подчиняется t-распределению с $(n-m-1)$ степенями свободы, где m — число независимых переменных. Результат компьютерного расчёта включает значения каждого b_i и соответствующих статистик с достигнутым уровнем значимости P. Иногда для β_i включается 95% доверительный интервал; если нет, то его можно рассчитать как $b_i \pm t_{0,05}\mathrm{SE}(b_i)$.

ВЫБРОСЫ И ВЛИЯТЕЛЬНЫЕ НАБЛЮДЕНИЯ

Как обсуждалось кратко в главе 28, выброс (наблюдение, которое противоречит большинству значений в наборе данных) может или не может быть влиятель-

ным (т.е. сильно изменять оценки параметров модели при своём исключении из расчёта). Выброс и/или влиятельное наблюдение могут иметь одно или оба следующих свойства.

- Большой остаток (остаток — разность между наблюдаемым и предсказанным значением зависимой (результирующей) переменной результата y, для его индивидуальных значений независимых переменных).
- Большая удалённость, когда конкретное значение независимой переменной x (или нескольких переменных x) находится на большом удалении от среднего значения. Удалённость от среднего значения x может считаться большой, если их отношение больше $2(k+1)/n$, где k — число независимых переменных в модели и n — число наблюдений в выборке.

Различные методы доступны для исследования степени чувствительности, с которой воздействуют на оценки модели отдельные наблюдения выборки. Мы можем считать подозрительными те наблюдения, которые (1) имеют большие остатки, большую удалённость значений, большую величину расстояния Кука (обобщённая мера влияния отдельного наблюдения, включающая как величину остатка, так и величину удалённости), или (2) проверены специальными диагностическими графиками, в которых влиятельные наблюдения могут стать визуально отличимыми.

ПРИМЕР

В главе 28 мы изучали соотношение между САД и ростом у 100 детей. Известно, что рост и масса тела коррелируют положительно, поэтому мы провели анализ множественной регрессии для исследования влияния массы тела (кг), роста (см) и пола (0 — мальчик, 1 — девочка) на САД (мм рт.ст.) у этих детей. Допущения, лежащие в основе этого анализа, проиллюстрированы на рис. 29-1–29-4.

Типичный результат компьютерного анализа содержится в приложении В. Анализ таблицы дисперсий показывает, что по крайней мере одна из независимых переменных взаимосвязана с САД (F=14,95 с 3 и 96 степенями свободы соответственно в числителе и знаменателе, P=0,0001). Скорректированное значение R^2 равно 0,2972 и означает, что 29,7% дисперсии САД можно объяснить моделью, т.е. различием в весе, росте и половой принадлежности детей. Таким образом, это даёт гораздо более хорошее согласие с данными, чем парная линейная регрессия в главе 28, где R^2=0,11. Типичный результат компьютерного анализа содержит следующую информацию о независимых переменных в этой модели.

Таблица 29-1. Результаты изучения соотношения между системным артериальным давлением и ростом

Переменная	Оценка параметра	Стандартная ошибка	95% доверительный интервал	t-критерий Стьюдента	Достигнутый уровень значимости
Свободный член	79,4395	17,1182	От 45,89 до 112,99	4,6406	0,0001
Рост	−0,0310	0,1717	От −0,37 до 0,31	−0,1807	0,8570
Масса тела	1,1795	0,2614	От 0,67 до 1,69	4,5123	0,0001
Пол	4,2295	1,6105	От 1,07 до 7,39	2,6261	0,0101

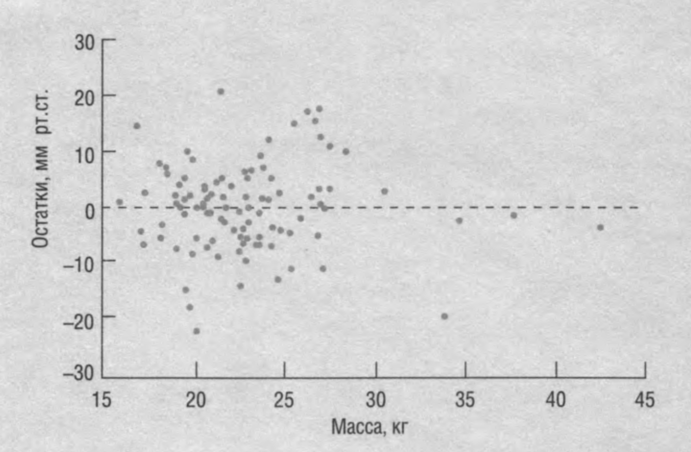


Рис. 29-1. Нет систематической структуры остатков в зависимости от веса. (Отметьте, что подобно рис. 28.2 диаграмма остатков из этой модели в зависимости от веса также не показывает систематической структуры).

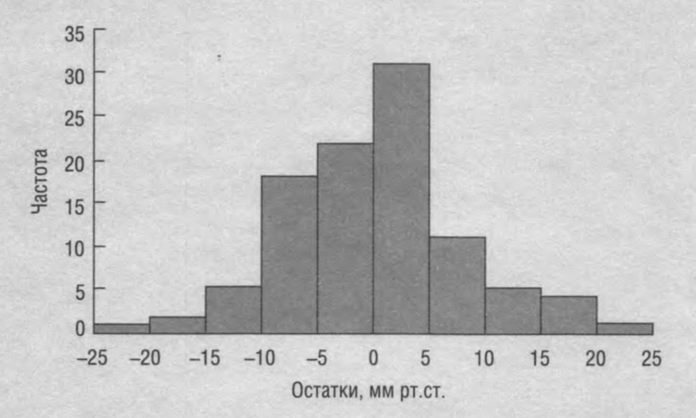


Рис. 29-2. Распределение остатков приблизительно нормально и дисперсия немного меньше, чем дисперсия от модели парной регрессии, отражая улучшенное предсказание в модели множественной регрессии по сравнению с парной моделью.



Рис. 29-3. Как и при модели парной регрессии, нет тенденции для остатков систематически увеличиваться или уменьшаться с предсказанными величинами. Следовательно, предположение постоянной дисперсии удовлетворено.

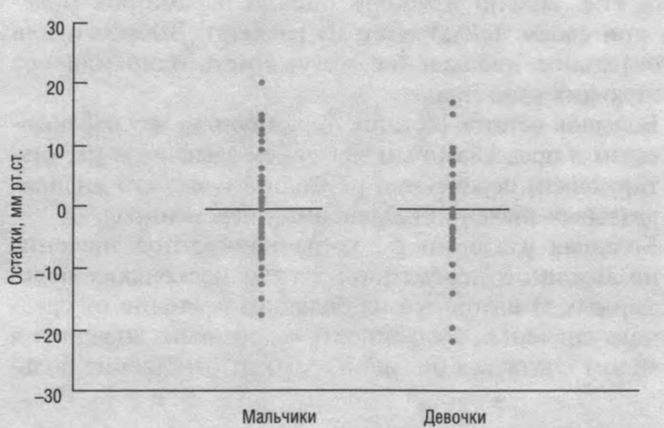


Рис. 29-4. Распределение остатков подобно у девочек и мальчиков, предполагая, что модель одинаково хорошо согласуется в обеих группах.

Уравнение множественной регрессии задаётся выражением:

$$САД=79,44-(0,03\times рост)+(1,18\times вес)+(4,23\times пол).$$

Соотношение между весом и САД высоко значимо (P <0,0001), с увеличением веса на один килограмм связано среднее увеличение на 1,18 мм рт.ст. с учётом роста и пола. Однако после учёта массы тела и пола ребёнка взаимосвязь между ростом и САД становится статистически незначимой (P=0,86). Это наводит на мысль, что значимое соотношение между ростом и САД в анализе парной регрессии отражает тот факт, что более высокие дети тяжелее, чем низкие. Имеется статистически значимое соотношение (P=0,01) между полом и САД; САД у девочек в среднем имеет тенденцию быть выше на 4,23 мм рт.ст., чем у мальчиков, даже после учёта возможных различий в росте и массе тела. Следовательно, как масса тела, так и пол — независимые предикторы САД у детей.

Мы можем рассчитать ожидаемое САД для детей с конкретным ростом и массой тела. Если первый ребёнок, упомянутый в главе 28, рост которого 115 см, — девочка с массой тела 37 кг, сейчас она имеет прогнозируемое САД в 79,44−(0,03×115)+ +(1,18×37)+(4,23×1)=123,88 мм рт.ст. (выше, чем 101,48 мм рт.ст., спрогнозированное в главе 28); если второй ребёнок ростом 130 см — мальчик с массой тела 30 кг, сейчас он имеет предсказанное САД в 79,44−(0,03×130)+(1,18×30)+(4,23×0)=110,94 мм рт.ст. (меньше, чем 108,68 мм рт.ст., предсказанное в главе 28).

30 Бинарные исходы и логистическая регрессия

ОБОСНОВАНИЕ

Логистическая регрессия очень похожа на линейную; её применяют, когда есть интересующий нас бинарный исход (т.е. наличие/отсутствие симптома или субъекта, который имеет/не имеет заболевания) и ряд предикторов. Из уравнения логистической регрессии можно определить, какие предикторы влияют на исход, и, используя значения предикторов пациента, оценить вероятность того, что он/она будет иметь определённый исход.

Мы начинаем с создания бинарной переменной, чтобы представить эти два исхода (т.е. «имеет болезнь»=1, «не имеет болезни»=0). Однако мы не можем применить эти два значения как зависимую переменную в анализе линейной регрессии, поскольку предположение нормальности нарушено, и мы не можем интерпретировать предсказанные величины, которые не равны нулю или единице. Фактически вместо этого мы берём вероятность того, что субъект классифицируется в ближайшую категорию (т.е. «имеет болезнь») зависимой переменной, и чтобы преодолеть математические трудности, применяем логистическое, или логит-преобразование (см. главу 9), в уравнении регрессии. Логит этой вероятности — натуральный логарифм (т.е. по основанию e) отношения вероятности «болезни» к вероятности «нет болезни», т.е.

$$logit(p) = \ln\frac{p}{1-p}.$$

УРАВНЕНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

Итеративный процесс, называемый методом максимального правдоподобия, а не обычная регрессия наименьших квадратов (так как мы не можем применить процедуру линейной регрессии) создаёт из данных выборки оценку уравнения логистической регрессии вида:

$$logit\ (p) = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k,$$

- x_i — i-я независимая переменная (i=1, 2, .., k);
- $logit\ (p)$ — оценка значения истинной вероятности того, что пациент с индивидуальным набором значений для x_1, .., x_k имеет заболевание;
- a — оценка константы (свободный член, пересечение);
- b_1, b_2, ... , b_k — оценки коэффициентов логистической регрессии.

Мы интерпретируем экспоненту отдельного коэффициента, например, e^{b_1}, как отношение шансов. Для индивидуальной величины x_1 — это оценка отношения шансов заболевания для (x_1+1) относительно оценённых шансов заболевания для x_1, при фиксированных значениях всех других предикторов x в уравнении. Если соотношение шансов равно единице, то эти два шанса равны. Значение отношения шансов больше единицы означает увеличенные шансы наличия заболевания, а значение меньше единицы — уменьшенные шансы наличия заболевания, когда x_1 увеличивается на одну единицу. Если заболевание встречается редко, соотношение шансов можно интерпретировать как относительный риск.

Мы можем манипулировать уравнением логистической регрессии, чтобы оценить вероятность того, что субъект имеет заболевание. Для каждого пациента с индивидуальным набором значений предикторов x_1, .., x_k мы рассчитываем величину z:

$$z = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k.$$

Тогда вероятность того, что субъект имеет заболевание, оценивается как:

$$p = \frac{e^z}{1+e^z}.$$

Поскольку модель логистической регрессии подогнана для логарифмической шкалы, эффекты x_i мультипликативны для шансов заболевания. Это означает, что их комбинированный эффект — произведение их раздельных эффектов (см. пример). Это не похоже на ситуацию для линейной регрессии, где эффекты x_i на зависимую переменную суммируются (аддитивны).

РЕЗУЛЬТАТ КОМПЬЮТЕРНОГО АНАЛИЗА

Для каждой независимой переменной

Исчерпывающий вывод компьютерного результата для анализа логистической регрессии включает для каждого предиктора оценку коэффициента логистической регрессии с его стандартной ошибкой, оценку отношения шансов (т.е. показательную функцию коэффициента) с доверительным интервалом для его истинной величины, статистику Вальда (тестирование нулевой гипотезы, что коэффициент логистической регрессии равен нулю, что эквивалентно тестированию гипотезы о том, что отношение шансов «заболевания», связанное с этой переменной, равно единице) и связанную с этим P-величину (достигнутый уровень статистической значимости). Мы используем эту информацию для того, чтобы определить, влияет ли каждая переменная на интересующую нас зависимую переменную (т.е. наличие или отсутствие заболевания), и количественно определить степень этого влияния. Можно использовать процедуры автоматического выбора предикторов, (см. главу 33) как при множественной линейной регрессии, чтобы выбрать лучшую комбинацию независимых переменных[1].

Правило «большого пальца» для максимального количества предикторов с целью включения в модель заключается в том, что максимальное число предик-

[1] Как и в случае множественной регрессии, для логистической регрессии также возможно вычисление стандартизованных коэффициентов регрессии, ранжируя которые возможно упорядочить предикторы по степени своего влияния на зависимую переменную. Прим. переводчика.

торов, включённых в модель, не должно быть больше, чем число значений одной из градаций (исходов), делённое на 10[1].

Чтобы оценить адекватность модели

Обычно наш интерес фокусируется на проверке независимых переменных в их влиянии на исход. Эта информация есть во всех современных пакетах статистических программ. Однако имеются расхождения между пакетами в способах оценки адекватности модели и её описаниях. Ваш компьютерный выход может содержать следующее (в том или другом виде).

- Величина, называемая $-2\times\log$ (правдоподобия), статистика отношения правдоподобия (LRS) или отклонение, оно имеет приблизительное распределение Хи-квадрат и показывает, сколь недостаточно модель соотносится со всеми предикторами (уровень значимости более критического означает плохое предсказание).
- Модель Хи-квадрат или Хи-квадрат для ковариат: она тестирует нулевую гипотезу о том, что коэффициенты регрессии в модели равны нулю. Значимый результат предполагает, что по крайней мере одна ковариата статистически значимо связана с зависимой переменной.
- Процент наблюдений, точно предсказанных «больной» или «небольной», с помощью данной модели. Эта информация может быть в классификационной таблице[2].
- Гистограмма: она имеет предсказанные вероятности вдоль горизонтальной оси и использует символы (такие, как 1 и 0) для обозначения принадлежности пациента к группе («больные» или «не больные»). Хорошая модель будет делить символы на две группы, которые показывают лишь небольшое перекрытие, или перекрытия вообще нет. Индексы эффективности предсказания имеются не во всех компьютерных программах, но они могут включать ложноположительные и ложноотрицательные доли результатов предсказания, а также чувствительность и специфичность модели. Наш совет — для полной информации обратиться к более углублённым работам.

СРАВНЕНИЕ ОТНОШЕНИЯ ШАНСОВ И ОТНОСИТЕЛЬНОГО РИСКА

Хотя отношение шансов часто берётся как оценка относительного риска, оно даст подобную величину, только если исход редок. Там, где исход нередок, отношение шансов будет больше, чем относительный риск, если последний больше единицы, в противном случае будет меньше, чем относительный риск. Хотя отношение шансов интерпретируется не так просто, как относительный риск, оно всё-таки имеет привлекательные статистические свойства и, таким образом, часто более предпочтительно (и должно применяться в исследовании «случай–контроль», когда относительный риск нельзя оценить прямо).

МУЛЬТИНОМИАЛЬНАЯ И ПОРЯДКОВАЯ ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Мультиномиальная (также называемая полихотомической) и порядковая, ординальная логистическая регрессия — расширения логистической регрессии: мы применяем их, когда имеем категориальную (качественную) зависимую переменную более чем с двумя градациями (категориями). Когда зависимая переменная номинальна (к примеру, имеется одно из трёх заболеваний спины: грыжа поясничного диска, хроническая боль в нижней части спины или острая боль там же), мы применяем мультиномиальную логистическую регрессию. Когда градации зависимой переменной ординальны или упорядочены (т.е. лёгкая, умеренная или сильная боль), мы применяем порядковую логистическую регрессию. Эти методы сложны и поэтому нужно обратиться к более углублённым работам и/или попросить совета специалиста, если вы хотите их использовать. Как простую альтернативу можно скомбинировать эти градации любым подходящим образом, чтобы создать новую бинарную переменную исхода, а затем провести обычный анализ бинарной логистической регрессии (признавая, что этот подход может быть неэкономичным по полученным результатам). Решение о том, как комбинировать эти категории, нужно принимать заранее, до того как смотреть данные, чтобы избежать смещения.

УСЛОВНАЯ ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Мы можем применять условную логистическую регрессию, когда имеем согласованные наблюдения (как в согласованном исследовании «случай–контроль») и хотим скорректировать возможные сопутствующие факторы. Анализ согласованного исследования «случай–контроль» с применением обычной логистической регрессии или методов, описанных в главе 16, не эффективен и не имеет достаточной мощности, поскольку наблюдения случая и контроля связаны друг с другом. Условная логистическая регрессия позволяет нам сравнивать случаи с контролем в том же согласованном наборе (т.е. каждую пару в случае согласования один к одному). В этой ситуации «исход» определяется пациентом, будучи случаем (обычно закодированным 1) или контролем (обычно закодированным 0). В этом случае продвинутые статистические пакеты могут иногда разрешить непосредственно применять условную логистическую регрессию, однако может возникнуть необходимость применить и регрессионную модель пропорциональных рисков Кокса.

[1] Peduzzi P., Concato J., Kemper E., Holford T.R., Feinstein A.R. A simulation study of the number of events per variable in logistic regression analysis // Journal of Clinical Epidemiology. — 1996. — Vol. 49. — P. 1373–1379.

[2] Пример такой таблицы дан в статье Н.П. Гарганеевой, В.П. Леонова «Логистическая регрессия в анализе связи артериальной гипертонии и психических расстройств» [42]. В качестве показателя степени согласия фактической и предсказанной принадлежности наблюдений (пациентов) к отдельным градациям зависимой переменной используют процент конкордации (согласия). Чем ближе он к 100%, тем лучше предикторы объясняют свою взаимосвязь с зависимой переменной. При использовании пошаговых алгоритмов отбора предикторов возможно в динамике проследить изменения в структуре такой таблицы. Прим. переводчика.

ПРИМЕР

В исследовании взаимосвязи между инфицированием вирусом герпеса человека типа 8 (HHV-8) и сексуальным поведением опросили 271 гомо/бисексуального мужчину относительно анамнеза по ряду передающихся половым путём болезней (гонорея, сифилис, простой герпес типа 2 [HSV-2] и ВИЧ). В главе 24 мы показали, что мужчины с анамнезом гонореи имели более высокую сероспространённость HHV-8, чем мужчины без такого анамнеза. Анализ множественной логистической регрессии был проведён для того, чтобы исследовать, был ли этот эффект просто отражением соотношений между HHV-8 и другими инфекциями и/или возрастом человека. Независимые переменные показывали наличие либо отсутствие каждой из четырех инфекций, причём каждая была закодирована как «0», если пациент не имел анамнеза какой-либо инфекции, или «1», если он имел эту инфекцию в анамнезе и возраст пациента в годах.

Типичный результат компьютерного анализа отражён в приложении В. Он показывает, что Хи-квадрат для предикторов равен 24,598 при 5 степенях свободы (P=0,0002), это означает, что по крайней мере один из предикторов связан с серостатусом HHV-8. Нижеприведённая таблица суммирует информацию о каждой переменной в модели.

Таблица 30-1. Исследование взаимосвязи между инфицированием вирусом герпеса человека типа 8 (HHV-8) и сексуальным поведением

Переменная	Оценка параметра	Стандартная ошибка	Хи-квадрат Вальда	Достигнутый уровень значимости	Отношение шансов	95% доверительный интервал для отношения шансов
Пересечение	−2,2242	0,6512	11,6670	0,0006	–	–
Гонорея	0,5093	0,4363	1,3626	0,2431	1,664	(0,71–3,91)
Сифилис	1,1924	0,7111	2,8122	0,0935	3,295	(0,82–13,28)
HSV-2-positivity	0,7910	0,3871	4,1753	0,0410	2,206	(1,03–4,71)
ВИЧ	1,6357	0,6028	7,3625	0,0067	5,133	(1,57–16,73)
Возраст	0,0062	0,0204	0,0911	0,7628	1,006	(0,97–1,05)

Эти результаты означают, что положительность HSV-2 (P=0,04) и статуса ВИЧ (P=0,007) независимо связаны с инфекцией HHV-8; пациенты, которые положительны по HSV-2, имеют риск быть HHV-8 серположительными в 2,21 раз больше (=exp[0,7910]), чем пациенты, сероотрицательные по HSV-2 после учёта других инфекций. Другими словами, риск сероположительности HHV-8 у этих пациентов повышается на 121%[1].

Верхний предел доверительного интервала для этих отношений шансов показывает, что этот повышенный риск может составлять 371% (4,71−1,0=371%). Инфекция HSV-2 — многократно задокументированный маркёр сексуальной активности. Таким образом, поскольку причина инфекции HHV-8 не HSV-2, эта связь может быть отражением сексуальной активности пациента.

Кроме того, имеется тенденция для сифилиса в анамнезе быть связанным с серостатусом HHV-8. Хотя это значимо в малой степени (P=0,09), мы должны отметить, что доверительный интервал включает достаточно высокое значение для отношения шансов, равное 13,28. Наоборот, нет указания на независимую связь между гонореей в анамнезе и сероположительностью HHV-8. Это наводит на мысль, что эта переменная, судя по критерию Хи-квадрат Пирсона (см. главу 24), связана с серостатусом HHV-8 из-за того факта, что многие мужчины с одной из других болезней в анамнезе, передающихся половым путём, в прошлом также имели гонорею. Нет статистически значимой связи между сероположительностью HHV-8 и возрастом; отношение шансов означает, что риск сероположительности HHV-8 увеличивается на 0,6% для каждого года жизни.

Вероятность, что 51-летний человек имеет HHV-8 инфекцию, если он имеет гонорею и позитивную реакцию на HSV-2 (но не имеет сифилиса и нет позитивной реакции на ВИЧ), оценивается как p=0,35, т.е.:

$$p=\{\exp(-0{,}6077)/\{1+\exp(-0{,}6077)\},$$

где −0,6077=0,2242+0,5093+0,7910+(0,0062×51).

[1] Приведённый пример вряд ли можно назвать удачным, так как из 6 коэффициентов (вместе со свободным членом) лишь два предиктора имеют достигнутый уровень значимости менее 5%, поэтому если использовать только статистически значимые коэффициенты, то вместо −0,6077 (показатель для экспоненты) мы получим −1,4332. Подставляя это значение в выражение для вычисления вероятности p, получаем вместо p=0,35 новое значение p=0,193. Вообще, более разумно в такой ситуации заново оценить параметры уравнения регрессии, используя, например, пошаговые алгоритмы отбора предикторов. Прим. переводчика.

31 Интенсивности и пуассоновская регрессия

ИНТЕНСИВНОСТИ

При любом длительном исследовании, изучающем появление события (например, такого, как смерть), мы должны принимать во внимание тот факт, что за всеми субъектами обычно наблюдают различные промежутки времени. Это может быть потому, что некоторые субъекты выпадают из исследования, или потому, что они включаются в него в разное время, и поэтому время длительного наблюдения различных людей может сильно варьировать. Поскольку пациенты с более длительным периодом наблюдения с большей вероятностью будут переживать (интересующее нас) событие, чем те, кто наблюдается в течение более короткого промежутка, мы считаем частоту, с которой произошло событие, на человека за период времени. Часто единица, представляющая удобный временной промежуток, — год (но это может быть минута, день, неделя и т.п.). Тогда частота события на человека в год (т.е. на человеко-год наблюдения) оценивается следующим образом:

Интенсивность = Общее количество случаев/Общее число лет наблюдения за всеми пациентами = Общее количество случаев/Человеко-лет наблюдения.

Каждое индивидуальное время наблюдения определяется обычно из времени, когда пациент вошёл в исследование, и до тех пор, когда наступило событие, или до тех пор, когда наблюдение завершилось, но событие не произошло. Общее время наблюдения — сумма всех индивидуальных времён наблюдения.

Частоту называют частотой заболеваемости (incidence), если событие — новый случай (например, болезнь), или частотой летального исхода, если событие — смерть. Когда частота очень маленькая, её умножают на удобный коэффициент, такой, как 1000, и повторно выражают как частоту на 1000 человеко-лет длительного наблюдения.

СВОЙСТВА ЧАСТОТЫ

- При расчёте частоты не делают различий между человеко-годами длительного наблюдения, которые происходят у одного и того же субъекта, и человеко-годами, которые происходят у разных субъектов. Например, человеко-годы наблюдения за 10 субъектами, каждый из которых наблюдался в течение 1 года, будут такими же, как человеко-годы наблюдения за одной персоной в течение 10 лет.
- Включать ли также множественные события от каждого субъекта (например, когда событие происходит более чем по одному поводу), зависит от интересующей нас гипотезы. Если мы заинтересованы только в первичных событиях, тогда наблюдение должно оканчиваться в точке, в которой субъект переживает свой первый случай, поскольку он больше не находится в группе риска для первичного события после этого времени. Когда в расчёт частоты включены множественные события от одного субъекта, мы имеем специальную форму сгруппированных данных и должны применять соответствующие статистические методы.
- Частота не может быть рассчитана в перекрёстном исследовании, поскольку этот тип исследования не включает время.

СРАВНЕНИЕ ИНТЕНСИВНОСТИ РИСКА

Риск события — просто общее число событий, делённое на число субъектов, включённых в исследование в его начале, без принятия во внимание длительности наблюдения. В результате риск события будет больше, когда субъекты наблюдаются дольше, поскольку они имеют больше возможности пережить событие. В противоположном случае интенсивность события в этих условиях должна оставаться относительно стабильной, поскольку эта интенсивность принимает во внимание длительность наблюдения.

ОТНОСИТЕЛЬНАЯ ИНЦИДЕНТНОСТЬ

Мы можем быть заинтересованными в сравнении интенсивности заболевания в группе субъектов, подвергнутых некоторому интересующему нас фактору ($\text{Интенсивность}_{\text{подверг.}}$), с интенсивностью в группе субъектов, не подвергнутых этому фактору ($\text{Интенсивность}_{\text{неподверг.}}$).

$$\text{Относительная инцидентность} = \text{Интенсивность}_{\text{подверг}} / \text{Интенсивность}_{\text{неподверг}}.$$

Относительная инцидентность (или отношение интенсивностей) интерпретируется аналогично относительному риску и отношению шансов (см. главы 16 и 30). Относительная инцидентность, равная 1, означает, что интенсивность заболевания в обеих группах одинакова, инцидентность более единицы означает, что интенсивность выше у субъектов, подвергшихся фактору, чем у субъектов, не подвергшихся, а относительная инцидентность менее единицы означает, что интенсивность ниже в группе, подвергшейся фактору.

Хотя относительная инцидентность часто берётся как оценка относительного риска, относительная инцидентность и относительный риск будут подобны, если событие (например, болезнь) случается редко. Когда событие происходит достаточно часто и субъекты наблюдаются различные промежутки времени, на интенсивность, а значит и на относительную инцидентность не будет влиять различная длительность наблюдения. Это не так для относительного риска, поскольку риск, и, следовательно, относительный риск, будут изменяться, так как субъекты наблюдаются более длительное время. Итак, относительная частота более предпочтительна тогда, когда в исследовании время длительного наблюдения варьирует между субъектами.

ПУАССОНОВСКАЯ РЕГРЕССИЯ

Что это такое?

Распределение Пуассона (названное по имени французского математика) — распределение вероятности редких событий, которые происходят случайно за временной интервал (или на пространстве) с постоянной средней частотой. Оно образует основу пуассоновской регрессии, которая применяется для анализа частоты какого-либо события (например, болезни), когда пациенты имеют различное время наблюдения. Она противопоставляется логистической регрессии (см. главу 30), которая связана только с тем, произошло или нет событие, и применяется для оценки соотношения шансов. При пуассоновской регрессии допускают, что интенсивность события среди субъектов с одинаковыми предикторами (например, возраст и пол) постоянна на протяжении всего периода исследования. В общем, мы хотим знать, какие предикторы влияют на интенсивность события, и хотим сравнить эту интенсивность в группах с различным воздействием фактора и/или предсказать эту интенсивность для групп пациентов с особыми характеристиками.

Уравнение и его интерпретация

Модель регрессии Пуассона принимает вид, очень похожий на модель логистической регрессии, причём каждая из них имеет (обычно) линейную комбинацию предикторов с правой стороны уравнения. Анализ регрессии Пуассона — зеркальное отражение анализа логистической регрессии в том, что мы преобразуем переменную исхода, чтобы обойти математические трудности. Мы применяем натуральное логарифмическое преобразование (ln) интенсивности и итеративный процесс (максимальное правдоподобие, чтобы создать оценку уравнения регрессии, исходя из данных выборки, вида:

$$\ln(r) = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k.$$

- x_i — i-я объясняющая переменная (i=1, 2, 3, .., k);
- r — оценка значения средней или ожидаемой интенсивности для субъекта с особой группой значений для x_1, .., x_k;
- a — оценка постоянного периода, обеспечивающего оценку логарифма пороговой интенсивности, когда все x_i-е в уравнении равны нулю (порог базовой интенсивности);
- b_1, b_2, .., b_k — оценки коэффициентов уравнения регрессии Пуассона. Экспонента отдельного коэффициента, например, e^{b_1} — оценка интенсивности заболевания, связанная с соответствующей переменной x_1. Для отдельной переменной x_i оценка интенсивности заболевания при изменении этой переменной на 1 равна относительной интенсивности, когда все остальные предикторы в уравнении неизменны. Если относительная интенсивность равна единице, тогда частота события та же, когда x_1 увеличивается на одну единицу. Величина относительной интенсивности выше единицы означает повышенную частоту события, а величина ниже единицы — пониженную частоту события, поскольку x_1 увеличивается на одну единицу. Как и при логистической регрессии, модели регрессии Пуассона располагаются на логарифмической шкале. Таким образом, действие предикторов x_i имеет мультипликативный (а не аддитивный, как в линейной множественной регрессии!)[1] эффект на частоту заболевания

Мы можем манипулировать уравнением пуассоновской регрессии для оценки интенсивности события для конкретного пациента с определённой комбинацией предикторных величин x_1, .., x_k. Для каждого набора значений предикторов x_1, .., x_k мы рассчитываем величину z следующим образом:

$$z = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k.$$

Тогда интенсивность появления события для этого пациента оценивается как e^z.

Применение отсечки

Хотя мы моделируем интенсивность, при которой происходит событие (например, число событий, делённое на число человеко-лет наблюдения), большинство статистических пакетов требует, чтобы вы определили как зависимую переменную число происходящих событий, а не интенсивность. Логарифм отдельных человеко-лет наблюдения пациента затем включается в модель как отсечка. Допуская, что мы заинтересованы только во включении отдельного события на человека, число событий, происходящих у каждого субъекта, примет величину 0 (если событие не произошло) или 1 (если событие произошло). Это даёт немного другую формулировку модели, которая разрешает обобщать оценки менее затратным для расчёта путём. Однако результаты от этой модели точно такие же, какими бы они были, если бы частоты были смоделированы.

Ввод данных для групп

Заметим, что когда все предикторы качественные, для упрощения процесса ввода данных можно использовать тот факт, что при расчёте частоты не различают человеко-годы наблюдения для одного индивидуума и для другого. Например, мы можем интересоваться влиянием только двух предикторов: полом (мужской или женский) и возрастом (младше 16, 16–20 и 21–25 лет) на частоту некоторого события. Тем самым эти два предиктора определяют шесть групп (мужчины моложе 16 лет, женщины моложе 16 лет... женщины в возрасте от 21 до 25 лет). Мы можем упростить ввод этих данных путём определения общего числа событий для всех субъектов внутри той же группы и всех человеко-лет наблюдения для этих субъектов. Тогда оценённая интенсивность в каждой группе рассчитывается как общее число событий, делённое на человеко-годы наблюдения в этой группе. Применяя этот подход, а не ввод данных для n пациентов один за другим, мы

[1] Аддитивный и мультипликативный характер взаимосвязи предикторов с зависимой переменной представляют собой два крайних случая более общего вида взаимодействия. Эти два взаимодействия объединяются с помощь преобразования Бокса–Кокса, которое было предложено авторами в 1964 г. Детали регрессии Бокса–Кокса и многих других типов регрессии прекрасно описаны в [54]. Прим. переводчика.

вводим данные для каждой из шести получившихся в результате групп, включая в модель фиктивные (индикаторные) бинарные переменные отдельно для пола и возраста. Отметьте, что при вводе данных таким образом невозможно приспособить цифровые ковариаты для определения групп или включить дополнительные ковариаты в модель, которая принимает различные величины для субъектов в группе.

Введение переменных, которые изменяются со временем

Путём деления периода длительного наблюдения на короткие интервалы можно ввести в модель переменные, которые со временем меняются. Например, мы хотим соотнести анамнез курения мужчин среднего возраста с вероятностью возникновения у них рака лёгких. За длительный период наблюдения многие из этих мужчин бросили курить, и в результате вероятность возникновения рака у них может понизиться. Таким образом, распределённые по категориям мужчины, согласно их статусу курения, могут дать слабое представление о воздействии курения на рак лёгких. Вместо этого мы делим каждый период длительного наблюдения мужчины на короткие временные интервалы таким образом, что их статус курения в каждом интервале остаётся постоянным. Далее проводим пуассоновскую регрессию, обрабатывая релевантную информацию в каждый короткий временной интервал для каждого мужчины (например, появление\не появление события, время его наблюдения и статус курения) как будто бы это «пришло» от другого мужчины.

Компьютерные результаты

Исчерпывающие результаты компьютерного анализа пуассоновской регрессии включают для каждого предиктора оценку коэффициента регрессии со стандартной ошибкой, оценку относительной интенсивности (например, экспоненту коэффициента) с доверительным интервалом для его истинного значения, статистику Вальда (статистику, проверяющую нулевую гипотезу, что данный коэффициент регрессии равен нулю или эквивалентно, что относительная интенсивность «болезни», связанной с этой переменной, равна единице) и соответствующую *P*-величину (достигнутый уровень статистической значимости). Как и в случае с исходом от логистической регрессии, мы можем оценить адекватность всей модели, используя $-2\times\log$ (отношения правдоподобия) (СМП или отклонение), и модель Хи-квадрат или Хи-квадрат для ковариат.

Пуассоновская экстрадисперсия

Одно опасение при подгонке модели пуассоновской регрессии — возможность появления экстрадисперсии, которая обычно означает сверхдисперсию. Это происходит, когда остаточная вариация больше, чем следует ожидать от модели Пуассона, возможно, потому что имеется посторонний человек (аномальное наблюдение, выброс) или потому что в модель не включён важный предиктор. Тогда стандартные ошибки обычно занижаются и, следовательно, доверительные интервалы для коэффициентов регрессии очень узкие, а *P*-величины очень малы. Способ исследовать наличие возможности сверхдисперсии — разделить величину $-2\times\log$ (правдоподобие), статистику отношения правдоподобия (СМП) или отклонение на степень свободы $(n-k)$, где n — число наблюдений в группе данных, а к — число параметров, введённых в модель (включая свободный член). Это частное должно быть приблизительно равно 1, если нет сверхдисперсии; величины значительно выше 1 могут означать наличие сверхдисперсии. Может также обнаружиться и недостаточная дисперсия, когда остаточная вариация меньше, чем следует ожидать от модели Пуассона, и где соотношение $-2\times\log$ (правдоподобие), делённое на $(n-k)$, значительно меньше единицы (например, если высокая встречаемость не может быть точно зарегистрирована). Уменьшенная дисперсия и сверхдисперсия могут вызывать опасения при выполнении логистической регрессии, где они называются экстрабиномиальной дисперсией.

Альтернатива пуассоновской регрессии

Когда группу пациентов наблюдают от естественной «начальной точки» (например, операции) до времени, когда у пациента развивается интересующий нас исход («конечная точка»), мы можем применить альтернативный подход, известный как анализ выживаемости, который в противоположность регрессии Пуассона не допускает, что эта «опасность» (интенсивность событий в небольшом интервале) была постоянна всё время. Этот подход детально описан в главе 44.

ПРИМЕР

Пациенты с ВИЧ, получающие высоко активное антиретровирусное лечение (HAART), обычно испытывают уменьшение ВИЧ вирусной нагрузки до уровней ниже лимита обнаружения пробы (первичный ответ). Однако после этой стадии у некоторых из этих пациентов могут обнаружить вирусологическую неудачу; это происходит, когда вирусная нагрузка пациента становится снова обнаружимой, хотя он ещё находится на лечении. Идентификация факторов, связанных с повышенной частотой вирусологической неудачи, может разрешить предпринять шаги для предупреждения этого появления. Имеется некоторое опасение, что частота вирусологической неудачи может увеличиваться с увеличением времени лечения. Поскольку пациенты наблюдаются на протяжении различного времени, для анализа подходит регрессионный анализ Пуассона.

516 пациентов, которые показали первичный ответ на лечение, были идентифицированы и прослежены до времени вирусологической неудачи или до их последней даты наблюдения, пока их вирусная нагрузка в это время оставалась подавленной. Наблюдение началось в первую дату, когда их вирусная нагрузка стала необнаружимой. Предиктором первичной значимости была длительность времени нахождения на лечении начиная с точки начального

ответа, но это была переменная, значения которой постоянно менялись для каждого пациента во время исследования, поэтому для того чтобы исследовать, изменялась ли частота вирусологической неудачи со временем, длительность времени нахождения на лечении с момента начального ответа была разделена на три временных интервала: менее 1 года, 1–2 года и более 2 лет (это породило 988 групп наблюдения), с приемлемым допущением, что частота вирусологической неудачи в каждый период была приблизительно постоянной. Частоты неудач в эти три временных периода затем сравнили.

Данные (длительность наблюдения в этом интервале, испытывалась или нет вирусологическая недостаточность в этом интервале, и релевантные предикторы) были занесены в матрицу наблюдений для каждого пациента в каждом интервале, в котором он длительно наблюдался. Рассматриваемые предикторы включали демографию, стадию болезни во время начала терапии, год начала HAART и получал или нет пациент лечение в прошлом.

Чтобы ограничить число ковариат в многомерной модели пуассоновской регрессии, для идентификации ковариат, связанных с вирусологической неудачей, применялась отдельная модель одновариантной регрессии Пуассона.

В целом за период в 718 человеко-лет 61 пациент имел вирусологическую неудачу, нескорректированная интенсивность этого события была равна 8,5 на 100 человеко-лет (95% ДИ): от 6,61 до 10,92. Интенсивность нескорректированной вирусологической неудачи была 8,13 (ДИ 6,31; 10,95) в первый год после первичного ответа на терапию, 12,22 (ДИ 7,33; 7,12) во второй год и 3,99 (ДИ 1,30; 9,31) в последующие годы. Результатом модели пуассоновской регрессии, которая включала только две фиктивных (индикаторных) переменных (см. главу 29), чтобы отразить категории 1–2 года и более 2 лет, причём каждая сравнивалась с градацией менее одного года, стало предположение, что начальный вирусологический ответ на терапию связан со временем, прошедшим после терапии, поскольку оно было статистически значимо связано с вирусологической неудачей (P=0,04). Кроме того, пол пациента (P=0,03), его/ее основная встречаемость CD8 (P=0,01) и состояние лечения ко времени начала текущего режима (ранее получаемое лечение, никогда не получаемое лечение, P=0,008) — все они были значимо связаны с вирусологической неудачей в одновариантных моделях пуассоновской регрессии. Таким образом, анализ многовариантной пуассоновской регрессии был проведён, чтобы оценить соотношение между вирусологической неудачей и длительностью терапии, после коррекции этого другими переменными. Результаты суммированы в табл. 31-1; полный компьютерный результат показан в приложении В.

Результаты этой многовариантной модели предполагали, что за период 1–2 года после первичного ответа был тренд к более высокой интенсивности вирусологической неудачи по сравнению с интенсивностью, наблюдаемой в первый год (интенсивность вирусологической неудачи уменьшилась на 53% за период 1–2 года), но более высокая интенсивность после второго года (интенсивность неудачи уменьшилась на 44% за этот период по сравнению с интенсивностью, наблюдаемой в первый год после первичного ответа), хотя ни один из этих эффектов не был статистически значимым. После коррекции для всех других переменных пациенты, получающие своё первое лечение, имели оценку интенсивности вирусологической неудачи, которая была на 44% ниже, чем эта интенсивность у пациентов, которые получали лечение ранее, оценка интенсивности вирусологической неудачи у мужчин была на 39% меньше, чем у женщин (однако это не было статистически значимо), а оценка интенсивности вирусологической неудачи была понижена на 65%, если исходная встречаемость CD8 была на 100 клеток/мм^3 выше.

См. также примеры в главах 32 и 33 для дополнительных анализов относительно этой модели Пуассона, включая оценки сверхдисперсии, добротность подгонки и линейность ковариат.

Таблица 31-1. Результаты многомерной пуассоновской регрессии для факторов, связанных с вирусологической неудачей

Переменная*	Оценка коэффициента	Стандартная ошибка	Оценка относительной инцидентности	95% доверительный интервал для относительной инцидентности	*P*-значение для статистики Вальда**
Время начала отклика на терапию (лет)					
<1	б/у***	–	1	–	–
1–2	0,4256	0,2702	1,53	0,90, 2,60	0,12
>2	–0,5835	0,4825	0,56	0,22, 1,44	0,23
Лечебный статус					
Перед лечением (0)	б/у	–	1	–	б/у
Без лечения (1)	–0,5871	0,2587	0,56	0,33, 0,92	0,02
Пол					
Женский (0)	б/у	–	1	–	–
Мужской (1)	–0,4868	0,2664	0,61	0,36, 1,04	0,07
CD8 (на 100 клеток/мм^3)	–1,0558	0,0267	0,35	0,33, 0,37	0,04

* Коды для бинарных переменных (пол и статус лечения) показаны в скобках. Время с начала начального ответа на терапию было включено путём включения ложных переменных, чтобы отразить периоды 1–2 года и более 2 лет после первичного ответа.
** Альтернативный метод оценки значимости категорийных переменных более чем с двумя градациями описан в главах 32 и 33.
*** б/у — базовый уровень для сравнения.
Адаптировано из работы Ms Colette Smith, Department of Primary Care and Population Sciences, Royal Free and University College Medical School, London, UK.

32 Обобщённые линейные модели

Статистическое моделирование включает использование простой и множественной линейной регрессии, логистической регрессии, пуассоновской регрессии и некоторых методов, которые имеют дело с данными по выживанию. Все эти методы основаны на генерации математической модели, которая наилучшим образом описывает зависимость между выходной переменной и предикторами. Генерация такой модели позволяет нам определить степень зависимости, в которой каждый из предикторов влияет на выходную переменную после учёта всех остальных предикторов модели, и, если необходимо, предсказать значение выхода для этого предиктора.

Обобщённая линейная модель (ОЛМ) может быть выражена формулой:

$$g(Y) = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k,$$

где Y — оценка предсказанного, среднего или ожидаемого значения зависимой переменной, которая подчиняется известному распределению вероятностей (нормальному, биноминальному, пуассоновскому); $g(Y)$, называемая функцией связи, — преобразование Y, которое даёт линейную зависимость от предикторов или объясняющих переменных x_1, .., x_k; b_1, .., b_k — оценки коэффициентов регрессии, соответствующие этим предикторам; a — постоянный член (свободный член).

Каждая из регрессионных моделей, описанных в предыдущих главах, может быть выражена как частный случай ОЛМ (см. табл. 32-1). Функция связи — логит отношения (т.е. логарифм шансов) в логистической регрессии и логарифм интенсивности в пуассоновской регрессии. Преобразование зависимой переменной необязательно в простой и множественной регрессии; функция связи определяется особенностями связи. После определения типа необходимой вам регрессии большинство статистических пакетов автоматически вводят в вычисления функцию связи без дополнительных спецификаций.

КАКОЙ ТИП МОДЕЛИ МЫ ВЫБИРАЕМ?

Выбор подходящей статистической модели будет зависеть от интересующего нас исхода (см. табл. 32-1). Например, если зависимая переменная — непрерывная числовая переменная, можно применить простую или множественную линейную регрессию для идентификации факторов, связанных с этой переменной. Если мы рассматриваем бинарный исход, (т.е. пациент умер или нет), а все пациенты наблюдаются одинаковое количество времени, тогда логистическая регрессия должна быть подходящим выбором модели. Отметьте, что можно выбрать различный тип модели, модифицируя формат зависимой переменной.

В частности, если есть непрерывный числовой результат, но одно или более допущений линейной регрессии не удовлетворяют критериям, можно выбрать классификацию переменной исхода на две группы, чтобы генерировать новую переменную бинарного исхода. Например, если зависимая переменная — систолическое кровяное давление (непрерывная числовая переменная) после шестимесячного периода антигипертензивной терапии, можно выбрать дихотомизацию САД как высокого или низкого, применяя конкретную точку деления, а затем применить логистическую регрессию для идентификации факторов, связанных с этим бинарным исходом.

При дихотомизации зависимой переменной, таким образом, мы можем упростить подгонку и интерпретацию статистической модели, но некоторая часть информации о зависимой переменной обычно теряется, поэтому всегда следует тщательно взвесить преимущества и недостатки этого подхода.

ПРАВДОПОДОБИЕ И ОЦЕНКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

При подгонке ОЛМ мы в целом применяем концепцию правдоподобия для того, чтобы оценить параметры модели. Для любой ОЛМ, характеризующейся известным распределением вероятности, группой предикторов и некоторыми потенциальными величинами для каждого из своих коэффициентов регрессии, правдоподобие модели (L) — вероятность того, что мы получим наблюдаемые результаты, если коэффициенты регрессии будут равны этим величинам. Оцениваем коэффициенты модели путём выбора величин для коэффициентов регрессии, которые максимизируют L (т.е. это величины, которые наиболее вероятно привели к нашим наблюдаемым результатам); процесс — оценка максимального правдоподобия (ОМП), а полученные при этом оценки параметров называются оценками максимального правдоподобия. ОМП — итерационный процесс, и поэтому требуется специальная компьютерная программа. Одно исключение для ОМП — в случае моделей простой и множественной линейной регрессии (с идентификацией функции взаимодействия), тогда мы обычно оцениваем параметры, применяя метод наименьших квадратов [оценки часто называются оценками обычных наименьших квадратов (ОНК)][1].

ОЦЕНКА АДЕКВАТНОСТИ ПОДГОНКИ

Хотя ОМП максимизирует L для данной группы предикторов, всегда можно далее улучшить L, включая дополнительные предикторы. При своей большой исключительности переопределённая модель — это уравнение регрессии, которое включает отдельную переменную для каждого наблюдения в группе данных.

Хотя такая модель должна прекрасно объяснять данные, применение её на практике ограничено, посколь-

[1] Помимо обычного МНК, существуют и иные разновидности: обобщённый, двушаговый, трёхшаговый, взвешенный и т.д. Прим. переводчика (глава 27); оценки ОНК и ОМП в этом случае идентичны.

Таблица 32-1. Выбор подходящих видов обобщённой линейной модели для применения с различными видами исхода

Тип исхода	Тип обычно используемой ОЛМ	См. главу
Непрерывный (количественный)	Парная или множественная линейная	28, 29
Бинарная		
Инцидентность заболевания в продольном исследовании (пациенты отслеживаются одинаковые периоды времени)	Логистическая	30
Бинарный исход в перекрёстном исследовании	Логистическая	30
Исследование с несвязанными наблюдениями «случай–контроль»	Логистическая	30
Исследование со связанными наблюдениями «случай–контроль»	Условная логистическая	30
Качественные исходы более чем с двумя градациями	Мультиномиальная или порядковая логистическая регрессия	30
Интенсивность событий или подсчёт событий	Пуассоновская регрессия	31
Время события*	Экспоненциальная, Вейбулла или Гомпертца модели	44

* Данные о наблюдениях во времени могут быть проанализированы регрессионной моделью пропорциональных рисков Кокса.

ку предсказание будущих наблюдений с помощью этой модели, по-видимому, будет плохим. Однако переопределённая модель всё-таки позволяет нам рассчитать величину L, которая должна быть получена, если удастся моделировать данные хорошо. Сравнение этой величины L с величиной, полученной после подгонки более простой моделью, с меньшим количеством переменных даёт способ оценки адекватности подгонки нашей модели. Мы рассматриваем отношение правдоподобия, отношение величины L, полученной из переопределённой модели, с величиной, полученной из подогнанной модели, чтобы сравнить эти две модели. Более точно рассчитывают СМП как:

$$\text{СМП} = -2\frac{\log(L_{\text{переопр}})}{\log(L_{\text{подогн}})} = $$

$$= -2\left\{\log(L_{\text{переопр}}) - \log(L_{\text{подогн}})\right\}.$$

СМП часто называют $-2\times\log$ (правдоподобия) или отклонение, которое приблизительно отвечает распределению Хи-квадрата со степенью свободы, равной разнице в числе параметров, использованных в этих двух моделях (т.е. $n-k$, где n — число наблюдений в выборке данных, а k — число параметров, включая свободный член (пересечение), в более простой модели). Нулевая гипотеза — гипотеза о том, что все дополнительные параметры переопределённой модели равны нулю; большая величина СМП даст значимый результат, показывая, что качество подгонки модели недостаточное.

СМП можно применять и в других ситуациях. В частности, для сравнения двух моделей, причём ни одна из них не является переопределённой, когда одна модель включает другую (т.е. большая модель включает все переменные, которые включены в меньшую модель, кроме дополнительных переменных). В этой ситуации тестовая статистика — разница между величиной СМП из модели с дополнительными переменными и из модели, не содержащей их. Тестовая статистика следует распределению Хи-квадрата со степенью свободы, равной числу дополнительных параметров, включённых в большую модель, и применяется для тестирования нулевой гипотезы о том, что все дополнительные параметры в большой модели равны нулю. СМП также можно применять для тестирования нулевой гипотезы о том, что все параметры, связанные с ковариатами модели, равны нулю, путём сравнения СМП модели, включающей ковариаты с той моделью, которая не содержит их. Она часто называется моделью Хи-квадрата или Хи-квадрата для ковариат.

ДИАГНОСТИКА РЕГРЕССИИ

При проведении любого вида регрессионного анализа важно рассматривать серийные диагностики регрессии. Они позволяют проверить подогнанные модели регрессии и поискать недостатки, которые могут повлиять на оценки параметров и их стандартные ошибки. В частности, необходимо рассмотреть, не нарушены ли допущения, лежащие в основе модели, и насколько воздействуют на наши результаты сильно влияющие наблюдения.

ПРИМЕР

В примере из главы 31 были применены тесты Вальда для идентификации индивидуальных факторов, связанных с вирусологическим откликом в группе из 516 ВИЧ-пациентов (с 988 группами наблюдения), которые принимали высокоактивное антивирусное лечение (HAART). В частности, мы интересовались, увеличилась ли частота вирусологической неудачи со временем после контроля других потенциально сопутствующих переменных, которые относились к вирусологической неудаче. Хотя результат начального интереса был бинарным (пациент испытывал или не испытывал вирусологическую неудачу), была выбрана модель регрессии Пуассона, а не логистическая модель, поскольку отдельные пациенты наблюдались различные периоды времени. Переменной исхода для проведённого анализа была интенсивность событий. В этой главе P-значения (уровни значимости) для переменных рассчитывали, применяя статистику отношения правдоподобия. В частности, чтобы рассчитать отдельные P-величины, связанные

с индикаторными переменными, представляющими собой время от первичного ответа на терапию, были подогнаны две модели. Первая модель включала переменные, относящиеся к статусу лечения (ранее полученное лечение, никогда не получаемое лечение), пол и исходную концентрацию CD8 (модель 1); вторая — эти же переменные, а также две индикаторные (модель 2). Затем рассматривались различия между величинами, полученными для −2×log (правдоподобие) (т.е. СМП или отклонение) (табл. 32-2). Полный компьютерный результат показан в приложении В.

Включение этих двух индикаторных переменных было связано с уменьшением величины −2×log правдоподобие) на 5,53 (=393,12−387,59). Эта тестовая статистика следует распределению Хи-квадрата с двумя степенями свободы (поскольку именно 2 дополнительных параметра были включены в большую модель); *P*-величина, связанная с этой тестовой статистикой, была равна 0,06, показывая, что индивидуальная связь между вирусологической неудачей и временем от первичного ответа не значима. Величина −2×log (правдоподобие) для модели 2 также позволила оценить адекватность подгонки этой модели путём сравнения её величины −2×log (правдоподобие) для распределения Хи-квадрата с 982 степенями свободы. Уровень значимости *P*, полученный для этого сравнения, был больше 0,99, что говорит о приемлемости подгонки модели. Однако следует отметить, что после включения в модель этих пяти переменных были некоторые данные об уменьшенной дисперсии, поскольку отношение −2×log (правдоподобие), делённое на её степени свободы, было равно 0,39, что гораздо меньше единицы. Это позволяет предположить, что количество остаточной дисперсии было меньше, чем следовало ожидать от модели Пуассона.

Таблица 32-2. Значения −2×log (правдоподобие), степени свободы и число параметров в моделях с включением и исключением времени первичного отклика на терапию

№	Переменные, включённые в модель	−2×log (правдоподобие)	Степени свободы для модели	Число параметров в модели, включая пересечение
1	Вид лечения, пол и исходный уровень CD8	393,12	984	4
2	Вид лечения, пол и исходный уровень CD8 и 2 фиктивных переменных для времени первичного отклика на терапию	387,59	982	6

33 Объясняющие переменные в статистических моделях

Какой бы вид статистической модели мы ни выбрали, нужно принять решение, какие предикторы включать в модель и каким способом они должны быть включены. Это решение будет зависеть как от вида предиктора (номинальный качественный, ординальный качественный или непрерывный числовой), так и от отношения между этими переменными и зависимой переменной.

НОМИНАЛЬНЫЕ ПРЕДИКТОРЫ

Обычно необходимо создать индикаторные переменные, чтобы исследовать эффект номинального качественного предиктора в регрессионном анализе. Отметим, что при оценке адекватности подгонки модели, которая содержит номинальную переменную более чем с двумя градациями, или значимости этой переменной важно включить все индикаторные переменные в модель одновременно. Если этого не сделать (т.е. включим только одну из индикаторных переменных для отдельного уровня качественной переменной), можно оценивать эффект этой переменной на исход лишь частично, поэтому предпочтительно оценить значимость этой переменной, применяя тестовую СМП, а не рассматривать отдельные *P*-величины для каждой из индикаторных переменных.

ОРДИНАЛЬНЫЕ ПРЕДИКТОРЫ

В ситуации, когда имеется ординальная переменная более чем с двумя градациями, можно применить один из двух следующих подходов.

- Рассматривать качественную переменную как непрерывное числовое измерение, «прикрепляя» числовое значение к каждой категории переменной. Этот подход разрешает полное применение распределения категорий, но он обычно допускает линейное отношение (когда числовые значения располагаются равномерно) между предиктором и зависимой переменной (или её преобразованием), и это должно быть подтверждено.
- Рассматривать качественную переменную как номинальный предиктор и создать для неё серию индикаторных переменных. Этот подход не принимает во внимание распределение категорий и поэтому требует много информации. Однако он не допускает линейного отношения с зависимой переменной и поэтому может быть предпочтительным.

Отличие величин СМП от этих двух моделей обеспечивает тестовую статистику для проверки линейного тренда (т.е. оценку того, даёт ли модель, допускающая линейное отношение, лучшую подгонку, чем модель, не допускающая линейного отношения). Эта тестовая статистика следует распределению Хи-квадрата со степенями свободы, равными разнице в числе параметров этих двух моделей; значимый результат предполагает нелинейность.

ЧИСЛОВЫЕ ПРЕДИКТОРЫ

Когда в модель включён числовой предиктор, оценка его коэффициента регрессии указывает на величину изменения зависимой переменной исхода при увеличении предиктора на одну единицу.

Таким образом, для простой и множественной линейной регрессии отношение между каждым предиктором и зависимой переменной допускается линейным. Для пуассоновской и логистической регрессий оценка коэффициента даёт меру воздействия увеличения предиктора на одну единицу на log зависимой переменной (т.е. модель предполагает линейное отношение между предиктором и log (интенсивности) или log (отношение шансов), кроме экспоненциального соотношения с действительной интенсивностью или отношением шансов). Важно проверить уместность предположения линейности (см. следующую главу) до включения предикторов в модели регрессии.

Оценка допущения (предположения) линейности

Чтобы проверить допущение линейности в модели простой или множественной регрессии, мы строим числовую зависимую переменную y против числового предиктора x или график остатков модели относительно x.

Исходные данные должны аппоксимировать прямую линию, а в остатках не должно быть визуально заметной тенденции. Можно оценить допущение линейности в логистической регрессии (см. главу 30) или пуассоновской регрессии путём разделения субъектов на категории с небольшим числом (5–10) равных подгрупп согласно их значениям x. При пуассоновской регрессии рассчитываем log (интенсивности) исхода в каждой подгруппе против средней точки диапазона величин для x для соответствующей подгруппы (см. рис. 33-1). Для логистической регрессии рассчитываем

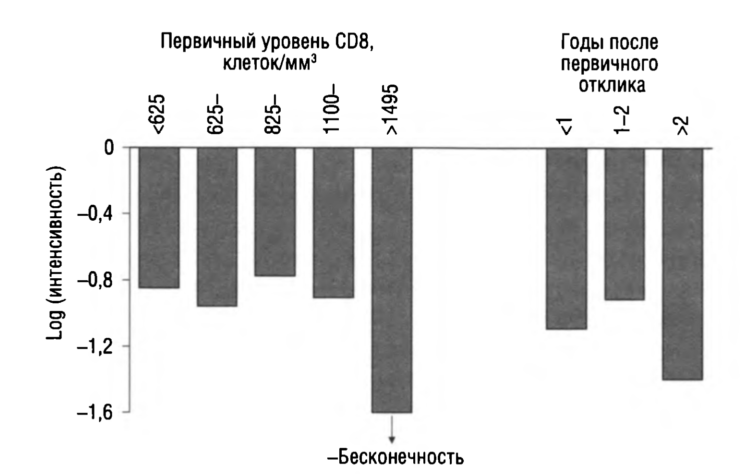


Рис. 33-1. График распределения значений log (интенсивность) в зависимости от исходного уровня CD8 и времени от первичного отклика при высокоактивном антиретровирусном лечении. Обе переменные на графике не обнаруживают линейной связи.

среднюю точку шансов для каждой подгруппы и строим график относительно этой точки. В каждом случае, если допущение линейности резонно, мы должны увидеть подобное ступенчатое увеличение (или уменьшение) интенсивности или шансов при перемещении между смежными категориями x.

Процедуры с нелинейностью. Если в какой-либо из этих диаграмм (графиков) обнаружена нелинейность, можно взять на вооружение ряд подходов.

- Замените x группой индикаторных переменных, созданных путём распределения субъектов на 3 или 4 подгруппы, согласно значению x (часто определяется, применяя терцили или квартили распределения). Эту группу индикаторных переменных можно ввести в модель многовариантной регрессии в качестве категорийных предикторов (см. пример).
- Преобразуйте переменную x каким-либо способом (например, логарифм или преобразование квадратного корня от x) так, чтобы получившееся в результате соотношение между преобразованной величиной x и зависимой переменной (или её логарифмом для пуассоновской регрессии, или её логит для логистической регрессии) было линейным.
- Найдите какое-либо алгебраическое описание, которое аппроксимирует нелинейное отношение, применяя более высокие порядки x (например, квадратичное или кубическое соотношение). Оно известно как полиномиальная регрессия. Вводим в уравнение эти предикторы, которые представляют собой соответственно более высокие порядки x. Так, например, если мы имеем кубическое отношение, наше оценённое уравнение множественной линейной регрессии имеет вид: $Y=a+b_1x+b_2x^2+b_3x^3$. Подгоняем эту модель и продолжаем работать с этим анализом точно таким же способом, как если бы квадратичный и кубический члены представляли собой обычные переменные (скажем, x_2 или x_3) в анализе множественной регрессии. Например, можно подогнать квадратичную модель, которая содержит предикторы «высота» и «высота»^{2}. Можно проверить линейность, сравнивая СМП в линейных и квадратичных моделях или тестируя коэффициент квадратичного члена.

ВЫБОР ПРЕДИКТОРОВ

Даже если модель не переопределённая, всегда есть опасность избыточной подгонки моделей при включении очень большого числа предикторов; это может вести к ложным результатам, не совместимым с ожиданиями, особенно если переменные тесно скоррелированы. Для модели множественной линейной регрессии обычное правило пальца — убедиться, что число наблюдений по крайней мере в десять раз больше, чем предикторов. Для логистической регрессии, для каждой из двух градаций исхода должно быть по крайней мере в десять раз больше наблюдений, чем число предикторов.

Часто большое число предикторов, как мы полагаем, могут относиться к зависимой переменной. Например, может оказаться, что многие факторы относятся к САД, включая возраст, диету и другие факторы образа жизни. Нужно включить в модель только те предикторы, которые с биологической или клинической точек зрения соотносятся с зависимой переменной. Можно исключить некоторые переменные, проведя одновариантный анализ (с менее строгим уровнем значимости, например, в 10, а не в 5%) для каждого предиктора, чтобы оценить, связан ли он с зависимой переменной, т.е. когда мы имеем числовую зависимую переменную, можем провести простой анализ регрессии, если этот предиктор числовой или непарный (t-тест, если он бинарный). Затем рассматриваем только те предикторы, которые были значимы на этой первой стадии для многомерной модели (см. пример в главе 31).

Процедуры автоматического выбора

Когда мы заинтересованы в применении модели для предсказания, а не в том, чтобы понять, влияет ли предиктор на исход, и не в оценке его эффекта, компьютерные процедуры интенсивного автоматического отбора обеспечивают средства идентификации оптимальной модели путём выбора некоторых этих переменных.

- Все подгруппы. Рассматривают каждую комбинацию предикторов; выбирают ту, которая даёт наилучшую подгонку, как описано моделью или СМП.
- Последовательное исключение. Включают все возможные переменные; те, о которых модель судит как о наименее важных (там, где это решение основано на изменении в R^2 или СМП), и далее последовательно на каждом шаге удаляют по одному предиктору до тех пор, пока нельзя будет удалить ни одну из оставшихся переменных, существенно не влияя на качество подгонки модели.
- Последовательное включение. Переменные, которые больше всего способствуют подгонке модели (на основе изменения в R^2 или СМП), последовательно включают в уравнение, пока никакая следующая переменная значительно не улучшит подгонку модели.
- Пошаговый отбор. Используют комбинацию процедур включения и исключения, которая начинается с последовательного включения, а затем в конце каждого «шага» проверяют процедурой исключения, чтобы убедиться, что все включённые переменные всё ещё необходимы.

Недостатки

Хотя эти процедуры избавляют от сложностей ручного выбора модели, они имеют некоторые недостатки.

- Возможно, что две или более моделей будут подходить к данным одинаково хорошо или эти изменения в группе данных будут создавать различные модели.
- Из-за множественного тестирования, которое происходит при повторном сравнении одной модели с другой внутри процедуры автоматического выбора, частота ошибки первого рода (см. главу 18) особенно высока. Таким образом, многие значительные находки могут обнаружиться случайно. Эту проблему можно уменьшить, выбирая более строгий уровень значимости (скажем, 1, а не 5%).

- Если модель повторно подгоняют к выборке данных, применяя, скажем, *m* переменных, остающихся в конечной модели автоматического отбора, оценки её параметров могут отличаться от оценок параметров модели автоматического выбора. Это потому, что процедура автоматического выбора применяет в своем анализе только те наблюдения, которые не имеют пропущенных значений по всем предикторам, предлагаемым для отбора, но размер выборки может быть больше, когда наблюдения включены, если они не имеют пропущенных величин только для релевантных *m* переменных.
- В результате получившиеся модели, хотя и математически подтверждаемые, могут не быть чувствительными. В частности, при включении серии индикаторных переменных, чтобы представить отдельную качественную переменную, автоматические модели могут включать только некоторые из индикаторных переменных, это приведёт к проблемам интерпретации модели.
- Для отбора лучшей модели подгонки нужно применять комбинацию этих процедур и полагаться на здравый смысл. Модели, которые сгенерированы с использованием автоматических процедур отбора, должны быть подтверждены, где это возможно, другими внешними наборами данных (см. «проверка устойчивости оценок»)[1].

ПРИМЕР

В главах 31 и 32 мы исследовали факторы, связанные с вирусологической неудачей у ВИЧ-пациентов, причём пациенты получали высокоактивное антиретровирусное лечение (HAART). В этом анализе с помощью многомерной пуассоновской регрессии исходная концентрация CD8 была включена как числовой предиктор (он делился на 100, так что каждая единица увеличения в значении переменной отражала увеличение концентрации на 100 клеток/мм^3). Результаты анализа показали, что более высокая концентрация CD8 была связана со значительно (статистически значимо) уменьшенной частотой вирусологической неудачи. Чтобы оценить устойчивость допущения линейности, связанного с этой переменной, было определено 5 групп на основе квинтилей распределения CD8, и в каждой из пяти групп была рассчитана частота неудачи. График log (интенсивность) в каждой из этих групп выявил, что соотношение было нелинейным, поскольку не было ступенчатой прогрессии (см. рис. 33-1). В частности, тогда как log (интенсивность) в четырёх нижних группах были сходны, в наиболее высокой группе (>1495 клеток/мм^3) никаких событий не произошло вообще, давая при этом значение минус бесконечность для log (интенсивность). По этой причине эти две верхних группы были скомбинированы для последующего анализа. Более того, было отмечено, что значительное число пациентов подлежало исключению из этого анализа, поскольку для них не было записи первичной концентрации CD8.

Таким образом, из-за отсутствия линейности между интенсивностью вирусологической неудачи и фактической концентрацией CD8 непрерывный предиктор, представляющий концентрацию CD8 в модели пуассоновской регрессии, был заменён на серию из четырёх индикаторных переменных. Пациенты с исходной концентрацией CD8 в диапазоне 825 <CDS< клеток/мм^3 обрабатывались как референтная группа для этих индикаторных переменных. Каждая из трёх индикаторных переменных дала сравнение одной из оставшихся CD8-групп с референтной группой, а 4-я индикаторная переменная дала сравнение с группой без концентрации CD8 с референтной группой. Результаты суммированы в табл. 33-1; полный компьютерный результат дан в приложении В. Сравнение величин для −2×log (правдоподобие) (т.е. СМП или отклонение), полученной от модели, которая включала эти четыре индикаторных переменных для концентрации CD8 (387,15) с такой же величиной, но полученной для модели, которая включала те же переменные отдельно от этих индикаторных переменных (392,50), дало значение уровня значимости P=0,25 (тестовая статистика 5,35 на четырёх степенях свободы). Итак, после её включения таким способом концентрация CD8 больше не имела статистически значимого влияния на вирусологическую неудачу в противоположность модели, которая неподходящим образом включала концентрацию CD8 как непрерывный предиктор. Однако отношение между вирусологической неудачей и статусом лечения, полом и временем, прошедшим с первичного ответа на терапию, осталось подобным.

[1] При этом важно понимать, что искомая модель необязательно должна быть единственной. Разумно иметь несколько подходящих моделей, которые хорошо интерпретируемы и дают высокое значение коэффициента детерминации R^2. С содержательной точки зрения предпочтительнее те модели, которые имеют меньше предикторов и более высокое значение R^2. Прим. переводчика.

Таблица 33-1. Результаты многомерной пуассоновской регрессии по факторам, связанным с вирусологической неудачей, после включения CD8 в модель в виде качественной переменной

Переменная*	Оценки коэффициентов	Стандартные ошибки коэффициентов	Оценки отношения интенсивностей	95% доверительный интервал для отношения интенсивностей	P-значение**(достигнутый уровень значимости)
Время первичного ответа терапии, годы					
<1	б/у***	–	1	–	–
1–2	0,4550	0,2715	1,38	0,93, 2,68	
>2	−0,5386	0,4849	0,58	0.23, 1,51	0,06
Статус лечения					
Ранее полученное лечение (0)	б/у	–	1	–	б/у
Никогда не получаемое лечение (1)	−0,5580	0,2600	0,57	0,34, 0,95	0,03
Пол					
Женщина (0)	б/у	–	1	–	–
Мужчина (l)	−0,4970	0,2675	0,61	0,36, 1,03	0,07
Концентрация CD8 (клеток/мм^3)					
<625	−0,2150	0,6221	0,81	0,24, 2,73	
≥625, <825	−0,3646	0,7648	0,63	0,16, 3,11	
≥825, <1100	б/у	–	1	–	–
≥1100	−0,3270	1,1595	0,78	0,07, 7,00	
неизвестно	−0,8264	0,6057	0,44	0,13, 1,43	0,25

* Коды для бинарных переменных (пол и статус лечения) показаны в скобках. Время от начального ответа на терапию было включено путём введения двух индикаторных переменных, чтобы отразить периоды 1–2 года и более 2 лет после начального ответа. Исходная концентрация CD8 была включена, как описано выше.

** Величины уровня значимости *P* получили, применяя СМП (см. главу 32), где индикаторные переменные применялись для включения более чем двух категорий переменной, величина *P* отражает комбинированный эффект этих индикаторов.

*** б/у — базовый уровень для сравнения.

34 Актуальные вопросы статистического моделирования

ВЗАИМОДЕЙСТВИЕ

Что это такое?

Статистическое взаимодействие (также известное как модификация эффекта) между двумя предикторами в анализе регрессии происходит, когда соотношение между одним из предикторов и зависимой переменной не одно и то же для различных уровней других предикторов, т.е. два предиктора не действуют независимо друг от друга на зависимую переменную. Например, предположим, что статус курения и алкогольный статус (каждый по отдельности) могут быть классифицированы на два уровня (курильщик/не курильщик и пьющий\не пьющий) и что каждый субъект принадлежит к одной категории каждой переменной. Если разница в диастолическом артериальном давлении (зависимая переменная) между курильщиками и не курильщиками больше в среднем у тех субъектов, кто не принимает алкоголь, чем у тех, кто его принимает, тогда мы говорим, что есть взаимодействие между курением и потреблением алкоголя.

Тестирование взаимодействия

Тестирование статистического взаимодействия в регрессионной модели обычно прямое, и многие статистические пакеты позволяют потребовать включения показателей взаимодействия. Если пакет не обеспечивает этой возможности, тогда элемент взаимодействия можно создать вручную, включая произведение релевантных переменных как дополнительный предиктор. Таким образом, чтобы получить величину переменной, которая представляет собой взаимодействие между двумя переменными (обе бинарные, обе числовые или одна бинарная и одна числовая), мы находим произведение этих двух переменных пациента. Если обе переменные числовые, интерпретация может быть легче, чем если мы создаём предиктор взаимодействия от двух бинарных переменных, полученный путём дихотомизации каждой числовой переменной. Если одна из двух переменных качественная более чем с двумя градациями, мы образуем от неё серию индикаторных переменных и применяем каждую из них вместе со второй интересующей нас бинарной или числовой переменной, чтобы генерировать серию предикторов взаимодействия. Эту процедуру можно расширить, если обе переменные качественные и каждая из них имеет более двух категорий. Показатели взаимодействия нужно включать в модель регрессии только после того, как включены главные эффекты (эффекты переменных без какого-либо взаимодействия)[1].

Отметьте, что статистический тест взаимодействия имеет обычно более низкую мощность. Это особенно следует иметь в виду, когда оба предиктора качественные и на каждую подгруппу, образованную путём сочетания каждого уровня одной переменной с каждым уровнем другой, приходится мало наблюдений, или эти подгруппы включают очень мало измеренных значений.

КОНФАУНДИНГ

Что это такое?

Сбивающая с толку (искажающая) переменная, или конфаундер, — это предиктор, который взаимосвязан в модели как с зависимой переменной, так и с одной или более предикторами. Например, мы можем интересоваться исследованием эффектов статуса курения и потребления алкоголя на частоту ишемической болезни сердца (ИБС) в когорте мужчин среднего возраста. Хотя известно, что и потребление алкоголя, и курение связаны с развитием ИБС, но эти две переменные также связаны друг с другом (т.е. более вероятно, что курят мужчины, которые потребляют алкоголь, чем непьющие). Любая модель регрессии, которая рассматривает эффекты влияния одного из предикторов на исход, но не включает конфаундер (т.е. модель, которая соотносит статус курения с частотой ИБС, не корректируя её на фактор употребления алкоголя), может неправильно представить истинную роль предиктора. Конфаундинг может или скрыть истинную связь, или искусственно создать ложную связь между предиктором и переменной исхода. Неудача в коррекции сбивающих с толку факторов будет вести к смещённым оценкам параметров модели.

Как обращаться с конфаундингом

С конфаундингом можно обращаться двумя способами.

- Создать подгруппы путём стратификации данных по уровням конфаундера (т.е. алкоголиков и не алкоголиков) и затем провести анализ в каждой группе отдельно. Этот подход простой и его можно рекомендовать, когда конфаундеров немного и объёмы подгрупп могут быть небольшими, и поэтому такой подход уменьшает мощность обнаружения значимого эффекта, а ложные значимые результаты могут возникнуть из-за множественного тестирования, если тестирование гипотез проводится в каждой группе. Могут также возникнуть затруднения в комбинации отдельных оценок интересующего нас эффекта для каждой подгруппы.
- Скорректировать конфаундеры в многовариантной модели регрессии. Этот подход, который особенно полезен, когда в исследовании много конфаундеров, даёт оценку соотношения между предиктором и зависимой переменной, которое нельзя объяснить соотношением между зависимой переменной и конфаундером.

---

[1] Современные статистические пакеты позволяют управлять составом анализируемых факторов, в том числе и последовательностью их включения в анализ. От этой последовательности может зависеть и величина уровня значимости *p*. Возможны ситуации, когда основные эффекты (факторы) оказываются статистически не значимыми, тогда как эффекты их взаимодействия значимы. Прим. переводчика.

7 Заказ 2051

Конфаундинг в нерандомизированных исследованиях

Мы должны быть особенно осторожными с конфаундингом при сравнении способов лечения в нерандомизированном клиническом когортном исследовании. В этом виде исследований характеристики субъектов могут быть неравномерно распределены в различных группах исследования. Например, субъекты могут быть отобраны для отдельного лечения на основе истории болезни, демографических факторов или факторов образа жизни. Некоторые из них могут быть связаны с исходами и поэтому будут смещать результаты лечения. Многомерные регрессионные модели можно применять для коррекции любых различий распределения факторов в различных группах лечения, но только если исследователи осведомлены о мешающих факторах и записали их в матрице данных. Рандомизированные контролируемые исследования редко страдают от конфаундинга, поскольку пациенты случайно распределены по группам лечения, и поэтому все ковариаты, как конфаундеры, так и другие предикторы, должны быть равномерным образом распределены в различных группах лечения.

КОРРЕКЦИЯ ДЛЯ ПРОМЕЖУТОЧНЫХ ПЕРЕМЕННЫХ

Когда известно, что переменная лежит на каузальном пути (в цепи причинно-следственных связей) между предиктором и интересующим нас исходом, она называется промежуточной переменной. Мы должны тщательно корректировать промежуточные переменные в многовариантных моделях. Рассмотрим ситуацию, когда проводится рандомизированное плацебо-контролируемое исследование эффекта нового липидснижающего препарата на частоту ИБС. Можно скорректировать любые расхождения между липидными уровнями пациентов в двух группах лечения на старте исследования (хотя это необязательно, если рандомизация проведена успешно), но нельзя корректировать изменения липидов, которые происходят в период исследования. Если скорректировать эти данные, нужно будет контролировать благоприятный эффект препарата, и таким образом любой эффект препарата, возможно, «исчезнет» (хотя это, быть может, покажет, в какой степени эффект препарата объясняется его воздействием на липидный уровень, что может быть полезным само по себе).

КОЛЛИНЕАРНОСТЬ

Когда два предиктора сильно закоррелированы, трудно оценить их индивидуальные эффекты в модели множественной регрессии. Как следствие, при том что каждая отдельная переменная может быть статистически значимо взаимосвязана с зависимой переменной в простейшей одномерной модели (т.е. когда имеется единственный предиктор), ни одна из них может не быть значимо связана с ней, когда оба предиктора включены во множественную модель. Эта коллинеарность (чаще называемая мультиколлинеарностью) может быть обнаружена путём проверки коэффициентов корреляции между каждой парой предикторов (часто отображённых в матрице корреляции) или визуальным впечатлением о стандартных ошибках коэффициентов регрессии в многовариантной модели; они могут быть существенно больше, чем коэффициенты в отдельных одномерных моделях, если есть коллинеарность. Самое простое решение этой проблемы — включить в модель только одну из таких переменных, хотя в ситуациях, где близко скоррелированы многие переменные, может появиться необходимость попросить статистического совета[1].

ПРОГНОСТИЧЕСКИЕ ИНДЕКСЫ И МЕТКИ РИСКА ДЛЯ БИНАРНОГО ОТВЕТА

При условии большого количества клинических характеристик пациента мы можем предсказать, разовьётся ли у него заболевание. Модели, которые часто подгоняются с применением регрессии пропорциональных рисков, логистической регрессии или другого метода, известного как дискриминантный анализ, могут применяться, чтобы идентифицировать факторы, которые статистически значимо связаны с исходом. Затем из коэффициентов этой модели может быть генерирован прогностический индекс, или метка риска, и величина этого индекса, рассчитанная для конкретного пациента, с целью оценки правдоподобия его заболевания. Однако модель, которая объясняет большую часть вариабельности в данных, не всегда подходит и для предсказания, у каких пациентов разовьётся болезнь, поэтому, поскольку мы вывели прогнозирующий индекс, основанный на модели, мы должны оценить устойчивость (валидность) этого индекса.

Устойчивость меток

Мы можем проверить устойчивость нашей метки несколькими способами.

- Создаём таблицу прогноза на основании матрицы данных, показывая число пациентов, у которых мы точно и неточно предсказываем статус болезни (подобно таблицам в главе 38). Величины, включающие чувствительность и специфичность, могут быть рассчитаны из этой таблицы.
- Распределяем по градациям пациентов согласно их меткам и рассматриваем частоту заболевания в различных градациях (см. пример); мы должны видеть соотношение между категориями и частотой заболевания, т.е. категории с большей меткой имеют большую частоту заболевания.

Ясно, что любая модель всегда будет хорошо работать на матрице данных, которая использовалась для генерации этой модели, поэтому, чтобы дать истинную оценку полезности метки, она должна быть обоснована на других, независимых, матрицах данных.

Там, где это невозможно, мы можем разделить исходные данные на две приблизительно одинаковые подгруппы. Первую подгруппу, известную как обучающая, применяют для генерации модели. Вторую, контрольную, группу применяют для валидизации результатов от обучающей выборки. Как следствие меньшего размера выборки в модель может быть включено меньшее число предикторов.

---

[1] При наличии мультиколлинеарности рекомендуется использовать пошаговые алгоритмы отбора предикторов. В этом случае из пары закоррелированных предикторов в уравнение регрессии будет отобран лишь один. Есть и другие алгоритмы оценки уравнения в таких условиях. Прим. переводчика.

Метод «складного ножа»

Джекнайфинг[1] — способ оценки параметров и обеспечения доверительных интервалов в несмещённой манере. Каждое наблюдение удаляют из исследования один раз за всё время, а остальные (n−1) наблюдения используют для оценки параметров модели. Этот процесс повторяют для каждого наблюдения в выборке, и получившиеся в результате оценки каждого параметра усредняют по всем n подвыборкам. Поскольку оценки, полученные таким способом, генерируются из многих различных подгрупп данных, он может быть валидизирован на основе полной матрицы данных, без использования подвыборки.

ПРИМЕР

Хотя между пациентами со СПИДом большие различия в прогнозе, о них часто думают как о единой однородной группе. Чтобы точно сгруппировать пациентов по их вероятному прогнозу, прогностическая оценка генерировалась на основе клинических наблюдений в одном центре Лондона за 363 пациентами, имеющими синдром приобретённого иммунодефицита (СПИД). За шестилетний период наблюдения умерло 159 (43,8%) пациентов.

Индекс (метка) был взвешенной суммой количества каждой группы тяжести (лёгкая, средняя или тяжёлая) болезни, определяющей СПИД, которую пережил пациент, и его минимального уровня клеток CD4 (измеренного в число клеток/мм^3) и был равен:

Индекс=300×число очень тяжёлых случаев СПИД (лимфома)
+100×число тяжёлых случаев СПИД (все случаи, кроме случаев, перечисленных как очень тяжёлые или лёгкие)
+20×число лёгких случаев СПИД (кандидоз пищевода, подкожная саркома Капоши, *Pneumocystis carinii* пневмония, экстралёгочный туберкулёз)
−1×минимальный уровень клеток CD4, измеренных со времени начала СПИДа.

С целью интерпретации этого индекса и для их обоснованности были идентифицированы три группы.

- СПИД I группы тяжести, индекс <0;
- СПИД II группы тяжести, индекс 0−99;
- СПИД III группы тяжести, индекс ≥100.

Обоснованность индекса была оценена путём рассмотрения частоты смерти (число смертей, делённое на общее число человеко-лет наблюдения) в каждой из групп.

Таблица 34-1. Индекс смертности в зависимости от групп (данные из первого центра в Лондоне)

Группа тяжести СПИД	Число смертельных исходов	Продолжительность жизни, до наступления смерти, человеко-лет	Индекс смертности
I	17	168,0	1,0
II	54	153,9	3,5
III	71	81,2	8,7

Таким образом, есть явный тренд к увеличению частоты смерти, по мере того как увеличивается показатель смертности. Оценка была также обоснована на группе пациентов из второго центра Лондона.

Таблица 34-2. Индекс смертности в зависимости от групп (данные из второго центра в Лондоне)

Группа тяжести СПИД	Число смертельных исходов	Продолжительность жизни, до наступления смерти, человеко-лет	Индекс смертности
I	65	828,5	0,8
II	229	579,6	4,0
III	322	361,3	8,9

Отмечены были удивительно схожие результаты, что подтверждает ценность этой системы индекса.

Адаптирована из Mocroft A.J., Johnson M.A., Sabin C,A. et al. Staging system for clinical AIDS patients. Lancet, 1995 — Vol. 346. — P. 12−17 (с разрешения издательства Elsevier).

[1] Метод предложил в 1949 г. М. Кенуй, название методу дал Дж. Тьюки. В 1969 г. М. Вайнцвайг назвал метод «скользящий контроль». Обобщением метода складного ножа является метод БУТСТРЕП. Прим. переводчика (метод «складного ножа») — метод оценки дисперсии и смещения оценивателя. Если объём выборки равен n, то оцениватель (алгоритм оценки) применяется к каждой подвыборке объёма n−1, полученной путём последовательного исключения одного измерения из анализа. Сумма квадратов разностей между каждой из полученных оценок и их средней, умноженная на (n−1)/n, — оценка дисперсии по методу складного ножа; разность между средней и первоначальной оценкой, умноженная на (n−1), — джекнайф-оценка смещения.

35 Проверка допущений

В ЧЁМ ПРИЧИНЫ ДЛЯ БЕСПОКОЙСТВА?

Компьютерный анализ даёт возможность работать с большими группами данных. Однако не делайте попытку их статистического анализа только потому, что они доступны вам на компьютерах. Обоснованность сделанных выводов покоится на соответствующем анализе, проведённом в любых данных обстоятельствах и требовании, что допущения, лежащие в его основе и присущие предложенному статистическому анализу, удовлетворены. Мы говорим, что анализ устойчив к нарушениям его допущений, если эти нарушения незначительно влияют на значение уровня значимости *P* и мощность критерия. Использование неустойчивого анализа может вести к неправильным заключениям.

ДАННЫЕ РАСПРЕДЕЛЕНЫ НОРМАЛЬНО?

Многие виды анализа имеют допущение о лежащем в его основе распределении данных. Следующие процедуры подтверждают приближённую нормальность, наиболее частую для допущений распределения.

Мы создаём точечный график (для небольших выборок) или гистограмму, график «стебель и листья» (см. рис. 4-2) или график «ящик с усами», чтобы показать распределение данных с эмпирической частотой. Заключаем, что это распределение приблизительно нормально, если оно имеет форму колокола и симметрично. Медиана в графике «ящик с усами» должна разделить пополам прямоугольник, образованный первой и третьей квартилями, и эти два уса (минимум и максимум) должны быть одинаковой длины, если данные распределены нормально.

Альтернативно мы можем создать нормальный график (предпочтительно на компьютере), который наносит распределение накопленной частоты данных (на горизонтальной оси) против частоты данных нормального распределения. Отсутствие нормальности показывается результирующим графиком, создающим кривую, которая, по-видимому, отклоняется от прямой линии (рис. 35-1). Хотя оба эти подхода субъективны, нормальный график более эффективен для меньших выборок. Для более объективной оценки нормальности следует использовать критерии Колмогорова–Смирнова[1] и Шапиро–Уилка[2], выполненные на компьютере.

РАВНЫ ЛИ ДВЕ ИЛИ БОЛЕЕ ДИСПЕРСИЙ?

Мы объясняли, как использовать *t*-критерий для сравнения двух средних и ANOVA для сравнения более чем двух средних. В основе этих анализов лежит допущение, что вариабельность наблюдений в каждой группе одинакова, т.е. мы требуем равенства дисперсий, описанных как гомогенность дисперсии или гомоскедастичность. Мы говорим о гетерогенности дисперсий, если они не равны.

- Можно использовать критерий Левене, применяя компьютерную программу, для проверки гомогенности вариации в двух или более группах. Преимущество критерия Левене в том, что он не сильно зависит от допущения нормальности. Критерий Бартлетта также можно использовать для сравнения более двух вариаций, но он не устойчив к отклонениям от нормальности.
- Можно использовать *F*-критерий (критерий отношения дисперсий) для сравнения двух дисперсий, при условии, что данные в каждой группе распределены приблизительно нормально (критерий неустойчив к нарушению этого допущения). Две оценки дисперсии — s_1^2 и s_2^2, вычисленные из n_1 и n_2 наблюдений соответственно. По умолчанию мы считаем s_1^2 большей из этих двух дисперсий, если они различаются.
- Мы также предполагаем однородность остатков в простой и множественной регрессии и в случайных моделях эффектов. Мы объяснили, как проверить это предположение в главах 28 и 29.

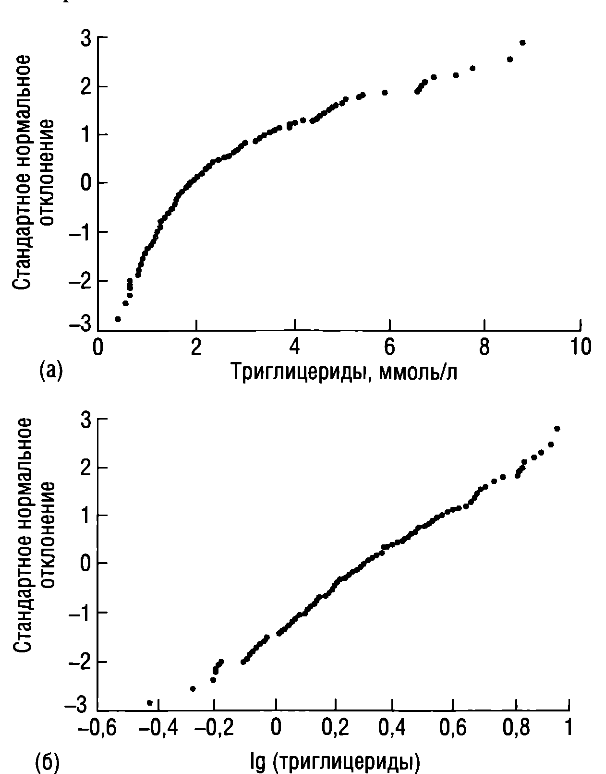


Рис. 35-1. (а) Нормальный график исходных уровней триглицеридов, описанный в главе 19. Они скошены (асимметричны), и результирующий нормальный график отчётливо показывает кривую линию. (б) Нормальный график логарифма уровня триглицеридов. Приблизительно прямая линия означает, что логарифмическое преобразование успешно устранило скошенность (асимметрию) данных.

[1] Более подробную информацию об этом критерии и детали его использования читатели могут найти в статье [87]. Прим. переводчика.

[2] W-критерий Шапиро–Уилка обладает максимальной мощностью при проверке гипотезы нормальности для малых выборок. К сожалению, в русскоязычной литературе он описан всего лишь в нескольких изданиях. Наиболее полно он описан в [162]. Прим. переводчика.

- ✧ 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✦ H_0: две популяционные дисперсии равны.
  - ✦ H_1: две популяционные дисперсии не равны.
- ✧ 2. Отобрать необходимые данные из выборки пациентов.
- ✧ 3. Вычислить значение статистики критерия, отвечающей H_0:

$$F = \frac{s_1^2}{s_2^2}.$$

  - ✦ Эта величина подчиняется F-распределению с числом степеней свободы (n_1-1) в числителе и (n_2-1) в знаменателе. Выбирая $s_1^2 \geq s_2^2$, мы гарантировали, что соотношение F всегда будет ≥ 1. Это позволяет использовать таблицы F-распределения, которые табулированы только для величин ≥ 1.
- ✧ 4. Сравнить величину статистики критерия с величинами из известного распределения вероятности. Обратиться к приложению А, табл. А5. Двусторонняя альтернативная гипотеза ведёт к двустороннему критерию Д.
- ✧ 5. Интерпретируйте величину P и результаты. Заметьте, что мы редко интересуемся дисперсиями самими по себе, так что мы обычно не вычисляем для них доверительные интервалы.

СООТНОСЯТСЯ ЛИ ПЕРЕМЕННЫЕ ЛИНЕЙНО?

Большинство методов, которые рассмотрены в главах 26–31 и описаны в главе 42, предполагают, что между двумя переменными есть линейные соотношения. Любые заключения, выведенные из этих анализов, полагаются на допущение о линейности, которое удовлетворено. Мы объяснили, как проверить линейность и как иметь дело с нелинейностью в регрессионном анализе в главах 28 и 29 (для простой и множественной регрессии) и в главе 33 (для других обобщённых линейных моделей, например логистической и пуассоновской регрессии).

ЧТО, ЕСЛИ ДОПУЩЕНИЯ НЕ ВЫПОЛНЯЮТСЯ?

- Продолжайте действовать по плану, признавая, что это может привести к неустойчивым результатам анализа. Осознавайте своё участие в том, что вы делаете. Не позволяйте себя (и других!) дурачить неуместным анализом, тем более что другие в похожих условиях уже делали это в прошлом!
- Возьмите подходящее преобразование исходных данных так, чтобы преобразованные данные удовлетворяли допущениям предложенного анализа. В регрессионном анализе это обычно означает преобразование независимой x переменной, хотя возможны и другие подходы.
- Если возможно, выполните непараметрический тест (см. главу 17), который исключает любые предположения о законе распределения данных (например, нормальность). Вы можете также встретить выражение «непараметрический регрессионный анализ»[1]. Его цель состоит в том, чтобы оценить функциональную форму (нежели параметры) взаимосвязи между зависимой переменной и одним или несколькими предикторами. Используя непараметрическую регрессию, мы ослабляем требования относительно линейности и подгоняем сглаженную кривую к данным так, что можем визуализировать тенденцию без того, чтобы запрашивать спецификацию параметрической модели.

ПРИМЕР

Рассмотрим пример непарного t-критерия из главы 21. Всего 98 детей школьного возраста были случайно выбраны на получение ингаляций беклометазона дипропионата или плацебо для определения их эффектов на одышку. Мы применили непарный t-критерий для сравнения среднего FEV1 в каждой группе в течение 16 мес, но требовались гарантии того, что лежащие в основе допущения (нормальность и равенство дисперсии) удовлетворены. Графики «стебель и листья» на рис. 4-2 показывают, что данные в каждой группе распределены приблизительно нормально. Чтобы проверить допущение равенства дисперсии в этих двух группах, мы использовали F-критерий.

- 1. Определить нулевую и альтернативную изучаемые гипотезы.
  - ✧ H_0: дисперсии измерения FEV1 в популяции детей школьного возраста одинаковы в обеих группах лечения.
  - ✧ H_1: дисперсии измерения FEV1 в популяции детей школьного возраста не одинаковы в обеих группах лечения.
- 2. Группа лечения: объём выборки n_1=50, стандартное отклонение s_1=0,29 л. Группа плацебо: размер выборки n_2=48, стандартное отклонение s_2=0,25 л.
- 3. Статистика критерия,

$$F = \frac{s_1^2}{s_2^2} = \frac{0{,}29^2}{0{,}25^2} = \frac{0{,}0841}{0{,}0625} = 1{,}336,$$

  что отвечает F-распределению с 50−1=49 и 48−1=47 степенями свободы соответственно в числителе и знаменателе.
- 4. Мы объясним F=1,34, обратившись к приложению А, табл. А5 для двустороннего критерия с 5% уровнем значимости. Поскольку табл. А5 ограничена степенями свободы 25 и 500 в числителе и 30 и 50 в знаменателе, мы вынуждены интерполировать. Требуемая табулированная величина при 5% уровне значимости лежит между 1,57 и 2,12; таким образом, P >0,05, поскольку 1,34 меньше, чем минимум этих двух величин (компьютерный расчёт даёт значение P=0,32).
- 5. Данных для того, чтобы отбросить нулевую гипотезу, что эти дисперсии равны, недостаточно. Уместно использовать непарный t-критерий, который предполагает нормальность и гомогенность дисперсии, чтобы сравнить средние величины FEV1 в этих двух группах.

[1] Eubank R.L. Nonparametric Regression and Spline Smooching, Marcel Dokker, 1999.

36 Расчёты размера выборки

ВАЖНОСТЬ РАЗМЕРА ВЫБОРКИ

Число пациентов в исследовании обычно ограничено этическими, временными и соображениями стоимости. Однако, если размер выборки слишком мал, мы не сможем обнаружить важный существующий эффект (т.е. мощность критерия будет неадекватной) и попусту растратим все ресурсы, поэтому вынуждены оптимизировать размер выборки, подводя баланс между размером выборки и факторами (такими, как мощность, величина эффекта лечения и уровень значимости), которые влияют на него. К сожалению, чтобы рассчитать требуемый размер выборки, должна быть некоторая идея относительно ожидаемых результатов[1].

ТРЕБОВАНИЯ

Объясним, как рассчитать оптимальный размер выборки в простых ситуациях; часто более сложные разработки могут быть упрощены с целью расчёта размера выборки. Если исследование включает ряд критериев, мы концентрируемся на наиболее важном или оцениваем объём выборки, требуемый для каждого критерия, и выбираем наибольший.

В центре внимания — вычисление оптимального размера выборки применительно к предложенной проверке гипотезы. Однако можно обосновать вычисление размера выборки на других аспектах исследования, например, на точности оценки или на ширине доверительного интервала (процессы, обычно принимаемые эквивалентными и не имеющие в исследованиях подчинённого положения).

Для того чтобы рассчитать объём выборки, требуется определить следующие величины ещё на этапе разработки исследования.

- Мощность — шанс обнаружения как статистически значимого, заданного эффекта, если он существует. Обычно мы выбираем мощность, равную 70–80% или более.
- Уровень значимости, α — граничный уровень, ниже которого мы будем отбрасывать нулевую гипотезу, т.е. это максимальная вероятность некорректно заключить, что эффект есть. Обычно мы фиксируем его как 0,05 или иногда 0,01 и отбрасываем нулевую гипотезу, если величина P меньше, чем эта величина.
- Вариабельность наблюдений, например, стандартное отклонение, если мы имеем числовую переменную.
- Наименьший интересующий эффект — величина эффекта, который клинически важен и который мы не хотим игнорировать. Это часто разность (например, разность в средних или пропорциях). Иногда она выражается в значениях стандартного отклонения наблюдений (стандартизованная разность).

Относительно легко выбрать мощность и уровень значимости критерия, который соответствует специфическим требованиям исследования. При условии особого клинического сценария можно уточнить эффект, который мы рассматриваем как клинически важный. Реальная трудность состоит в обеспечении оценки вариации числовой переменной до того, как мы отобрали данные. Эту информацию можно получить из других опубликованных исследований с аналогичными результатами или самостоятельно выполнить это экспериментальное исследование.

МЕТОДОЛОГИЯ

Рассчитать объём выборки можно несколькими способами, каждый из которых требует, по существу, той же информации (описанной в «Требованиях»), чтобы его выполнять.

- Общие формулы — они могут быть сложными, но необходимы в некоторых ситуациях (например, чтобы сохранить мощность в кластерном рандомизированном исследовании). Мы умножаем размер выборки, который был бы необходим, если мы выполняли индивидуальное рандомизацированное распределение, планируя эквивалентный эффект, на $[1+(m-1)\times\rho]$, где m — средний объём группы и ρ — величина внутригруппового коэффициента корреляции (см. главу 42).
- Быстрые формулы — они существуют для отдельных величин мощности и уровней значимости и некоторых критериев проверки гипотез (например, формула Лера[2]).
- Специальные таблицы[3].
- Номограмма Альтмана — лёгкая в применении диаграмма, которая подходит для различных критериев. Детали даны в следующей главе.
- Компьютерное программное обеспечение — его преимущество в том, что результаты можно представить графически или в таблицах, чтобы показать последовательность влияния изменения факторов (например, мощности, величины эффекта) на требуемый объём выборки.

НОМОГРАММА АЛЬТМАНА

Система обозначений

В табл. 36-1 показана система обозначений для применения номограммы Альтмана с целью оценки размера выборки из двух групп наблюдений одинакового

[1] В некоторых современных пакетах есть специальные процедуры оценки требуемого объёма наблюдений, которые обеспечивают необходимую мощность используемого статистического критерия для заданных условий исследования. Достаточно подробно процедуры оценки требуемого объёма наблюдений для разных методов описаны в изданиях, приведённых по адресам http://www.biometrica.tomsk.ru/razdel_2_1.htm — http://www.biometrica.tomsk.ru/razdel_2_12.htm Прим. переводчика.

[2] Lehr R. Sixteen s squared over d squared: a relation for crude-sample size estimates. Statistics in Medicine, 1992. — Vol. 11. — P. 1099–1102.

[3] Machin D., Campbell M.J. Statistical Tables for the Designof Clinical Trials — 2 nd edn. — Oxford: Blackwell Scientific Publications, 1995.

Таблица 36-1. Информация для применения номограммы Альтмана

Критерий проверки гипотез	Стандартизованная разница	Пояснение N в номограмме	Терминология
Непарный t-критерий (глава 21)	δ/σ	$N/2$ наблюдений в каждой группе	δ: наименьшая клинически значимая разность в средних. σ: принятое стандартное отклонение наблюдений, одинаковое в каждой из двух групп. Вы можете оценить его, применяя результаты подобного исследования, проведённого ранее, или опубликованную информацию. А также вы могли бы провести пилотное исследование для его оценки. Другой подход — выразить σ из множества стандартных отклонений (например, при проверке разности двух стандартных отклонений)
Парный t-критерий (глава 20)	$2\delta/\sigma_d$	N пар наблюдений	δ: наименьшая клинически значимая разность в средних. σ_d: стандартное отклонение разности количественного признака, обычно оценённое в пилотном исследовании
Хи-квадрат критерий (глава 24)	$\dfrac{p_1 - p_2}{\sqrt{\bar{p}(1-\bar{p})}}$	$N/2$ наблюдений в каждой группе	p_1-p_2: наименьшая клинически важная разность в долях (пропорциях) «успеха» в двух группах. Одна из этих пропорций обычно известна, а необходимая разность оценена путём рассмотрения, какую величину должна принять другая пропорция, чтобы составить заметное изменение. $\bar{p} = \dfrac{p_1 + p_2}{2}$

объёма для трёх часто применяемых критериев проверки гипотез о средних и пропорциях.

Метод

Для каждого критерия рассчитываем стандартизованную разность, откладываем её значение на левой оси номограммы (приложение В) и соединяем эту точку линией с осью мощности, которую мы определили на правой вертикальной оси. Требуемый объём выборки дан в точке пересечения линии, соединяющей левую и правую оси, с осью объёма выборки.

Заметьте, что мы можем также применить номограмму для оценки мощности критерия гипотезы для данного объёма выборки. Это полезно, если мы хотим узнать ретроспективно, можем ли мы отнести отсутствие значимости критерия проверки гипотезы к неадекватному размеру выборки. Также помните, что широкий доверительный интервал для интересующего нас эффекта означает плохую мощность.

БЫСТРЫЕ ФОРМУЛЫ

Для непарного t-критерия и критерия Хи-квадрат Пирсона мы можем применить формулу Лера, чтобы рассчитать объём выборки для мощности 80% и двустороннего уровня значимости в 0,05. Требуемый размер выборки в каждой группе равен 16/(стандартизованная разница).

Если стандартизованная разность небольшая, она переоценивает объём выборки. Заметьте, что числитель 21 (вместо 16) относится к мощности в 90%.

ФОРМУЛИРОВКА МОЩНОСТИ

Часто существенно важно и всегда полезно включать высказывание о мощности в статьи и протокол исследования и в главу «Методы», чтобы показать, что размер выборки был тщательно продуман ещё на стадии разработки исследования. Типичной формулировкой может быть такая: «Потребовалось 84 пациента в каждой группе, чтобы иметь 90% шанс обнаружения разности в средних в 2,5 дня (SD=5 дней) при 5% уровне значимости, применяя непарный t-критерий» (см. пример 1).

КОРРЕКТИРОВКА

Может потребоваться скорректировать объём выборки.

- Чтобы учесть последующие потери, вводя с самого начала большее количество пациентов в исследование. Если мы полагаем, что норма потерь наблюдений будет r%. Тогда скорректированный объём выборки будет получен умножением нескорректированного объёма на величину $100/(100-r)$.
- Чтобы иметь независимые группы различных размеров. Это желательно, когда одна группа ограничена в объёме, возможно, потому что данная болезнь редка в исследовании «случай–контроль» или потому что новое лечение, препарат может быть применим в небольших объёмах. Отметим, однако, что несбалансированность объёмов выборок обычно приводит к большому общему объёму наблюдений, когда проводится сравнение со сбалансированным проектом, если должен быть обеспечен аналогичный уровень мощности. Если отношение объёмов выборок в этих двух группах равно k (например, k=3, если мы требуем, чтобы одна группа была в три раза больше другой), скорректированный полный размер выборки равен:

$$N' = \frac{N(1+k)^2}{4k},$$

где N — нескорректированный полный объём наблюдений, рассчитанный для одинаковых по объёму групп. Тогда $N'/(1+k)$ пациентов будут в меньшей по объёму группе, и остающиеся пациенты будут в большей группе.

ПРИМЕР 1

Сравнение средних в независимых группах с применением непарного t-критерия.

Цель — проверить эффективность суспензии ацикловира (15 мг/кг) для лечения детей от 1 до 7 лет с герпетическим гингивостоматитом, длящимся менее 72 ч.

План исследования — рандомзированное двойное слепое плацебо-контролируемое исследование с «лечением», проводимым 5 раз в день в течение 7 дней.

Основной показатель результата для определения объёма выборки — длительность пероральных повреждений.

Вопрос относительно объёма выборки — сколько детей требуется, чтобы иметь 90% мощность обнаружения 2,5-дневной разницы в длительности пероральных повреждений между этими двумя группами при 5% уровне значимости? Авторы предполагают, что стандартное отклонение длительности пероральных повреждений составляет около 5 дней.

Применение номограммы:

$$\delta=2{,}5 \text{ дня и } \sigma=5 \text{ дней}.$$

Таким образом, стандартизованная разность равна:

$$\frac{\delta}{\sigma}=\frac{2{,}5}{5}=0{,}50.$$

Линия, соединяющая стандартизованную разность в 0,5 и мощность в 90%, пересекает ось объёма выборки приблизительно возле 160. Таким образом, требуется около 80 детей в каждой из групп. Отметьте: (i) если бы δ была увеличена до 3 дней и стандартизованная разность равнялась бы 0,6, то требуемый размер выборки уменьшился бы в целом приблизительно до 118, т.е. по 59 в каждой группе, и (ii), используя приведённую выше спецификацию, исследователь обнаружит, что при изучении лечения детей препаратом ацикловиром с группой плацебо (т.е. k=2) скорректированный объём выборки был бы равен:

$$N'=\frac{N(1+k)^2}{4k}=\frac{160(1+2)^2}{(4\times 2)}=180,$$

с 180/3=60 детьми на плацебо и оставшиеся 120 детей на ацикловир. Рис. 18-1 показывает кривую мощности критерия для этого примера.

Быстрая формула

Если мощность 90%, требуемый размер выборки в каждой группе:

$$\frac{21}{(\text{стандартизированная разность})^2}=\frac{21}{(0{,}50)^2}=84.$$

Amir J., Haral L., Smettana Z., Varsano I. Treatment of herpes simplex gingivostomatitis with aciclovir in children: a randomised double–blind placebo–controlled study // British Medical Journal, 1977. — Vol. 314. — P. 1800–1803.

ПРИМЕР 2

Сравнение двух пропорций в независимых группах с применением критерия Хи-квадрат Пирсона.

Цель — сравнить эффективность инъекций глюкокортикоида с физиотерапией для лечения болезненного ригидного плеча.

План исследования — РКИ, в котором пациенты случайным способом назначены на шестинедельное лечение, содержащее или максимум три инъекции, или двенадцать 30-минутных сеансов физиотерапии для каждого пациента.

Основной показатель результата для определения объёма выборки — лечение считается успешным после 7 нед, если пациент считает себя полностью выздоровевшим или у него наступило улучшение (по шеститочечной шкале Ликерта).

Вопрос относительно объёма выборки — сколько требуется пациентов, чтобы иметь 80% мощность обнаружения клинически важной разницы в частоте успеха в 25% между двумя группами при 5% уровне значимости? Авторы считают частоту успеха в 40% в группе, имеющей менее успешное лечение.

Применение номограммы: p_1=0,40, p_2=0,65, так что

$$\overline{p}=\frac{0{,}40+0{,}65}{2}=0{,}525,$$

поэтому стандартизованная разность равна:

$$\frac{p_1-p_2}{\sqrt{\overline{p}(1-\overline{p})}}=\frac{0{,}25}{\sqrt{0{,}525\times 0{,}475}}=0{,}50.$$

Линия, соединяющая стандартизованную разность в 0,5 и мощность в 80%, пересекает ось объёма выборки на 120, поэтому требуется 60 пациентов в каждой группе [отметьте: (i) если бы мощность была увеличена до 85%, требуемый размер выборки увеличился бы в целом примерно до 140, т.е. в каждой группе потребовалось бы по 70 пациентов, и (ii) если уровень потери наблюдений, как ожидалось, будет порядка 20%, то скорректированный объём выборки (при мощности критерия в 80%) был бы 120×100/(100−20)=150, с 75 пациентами в каждой группе]. На рис. 18-2 представлены кривые мощности критерия для этого примера.

Быстрая формула

Если мощность критерия 80%, требуемый размер выборки в каждой группе равен:

$$\frac{16}{(\text{стандартизированная разность})^2}=\frac{16}{(0{,}50)^2}=64.$$

van der Windt D.A.W.M., Koes B.W. Deville W., de Jong B.A., Bouter M. Effectiveness of corticosteroid injections with physiotherapy for treatment of painful shoulder in primary care: randomised trial // British Medical Journal, 1998. — Vol. 317. — P. 1292–1296.

37 Представление результатов[1]

ВВЕДЕНИЕ

Существенный аспект статистики — способность суммировать важные свойства анализа. Мы должны знать, какие данные включить и как отобразить результаты таким образом, чтобы облегчить другим получение уместной и важной информации и сделать точное заключение. Эта глава описывает ключевые свойства такого представления.

ЧИСЛОВЫЕ РЕЗУЛЬТАТЫ

- Давайте цифры только в той степени точности, которая подходит (как норма, на одну значимую цифру больше, чем исходные данные). Если анализ данных проведён вручную, округлять только в конце вычислений.
- Давайте число наблюдений, на которых основана любая сводка (например, процент).
- Описывайте любые аномальные значения и объясняйте, как они обработаны (см. главу 3).
- Включайте единицы измерения.
- Когда интерес фокусируется на параметре (например, средняя, коэффициент регрессии), всегда указывайте точность его оценки. Мы рекомендуем применять для этого доверительный интервал, но стандартная ошибка также допускается. Избегайте применения условного обозначения ± для «среднее ±SEM», поскольку, прибавляя и вычитая SEM, мы создаём 67% доверительный интервал, который может ввести в заблуждение тех, кто применял 95% доверительные интервалы. Лучше показать стандартную ошибку в скобках после оценки параметра [например, среднее =16,6 г (SEM 0,5 г)].
- Когда интерес фокусируется на распределении наблюдений, всегда обозначайте систему измерения «рассеяния» данных. Диапазон величин, который исключает аномальные значения (обычно диапазон величин, содержащий центральные 95% наблюдений, глава 6), — полезное описание. Если данные распределены нормально, этот диапазон аппроксимируется средним выборки ±1,96×стандартное отклонение. Вместо этого вы можете оценить среднее и стандартное отклонение [например, среднее =35,9 мм (СО 2,8 мм)], но тогда оценку диапазона вы оставляете читателю.

ТАБЛИЦЫ

- Не давайте слишком много информации в таблице.
- Давайте краткий, информативный и недвусмысленный заголовок.
- Помечайте каждый ряд и колонку.
- Помните, что легче просматривать информацию вертикально по колонке, чем горизонтально, вдоль строки.

[1] Дополнительную информацию по этому непростому вопросу читатели могут найти в статье [96], а также в книге [130]. Прим. переводчика.

ДИАГРАММЫ

- Делайте диаграмму простой и устраняйте ненужные «украшения» (например, создавая круговой график трёхмерным).
- Давайте краткий, информативный и недвусмысленный заголовок.
- Помечайте все оси, сегменты и решётки и объясняйте значение символов.
- Устраняйте мешающие результаты путём подчеркивания шкалы на оси.
- Указывайте, где два или более наблюдения лежат в одинаковом положении на двухмерном графике, например, применяя различные символы.
- Обеспечьте, чтобы вся уместная информация содержалась в диаграмме (например, соедините парные наблюдения).

ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ В СТАТЬЕ

Представляя результаты в статье, мы должны гарантировать, что она содержит достаточно информации для читателя, чтобы он понял, что сделано. Читатель должен иметь возможность воспроизвести результаты, при условии подходящей компьютерной программы и данных. Все аспекты разработки исследования и статистическая методология должны быть полностью описаны (см. также «Требования CONSORT»).

Результаты проверки гипотез

- Включайте относящуюся к делу диаграмму, если уместно.
- Обозначайте интересующую гипотезу.
- Назовите критерий и сформулируйте, какой он: одно- или двусторонний.
- Подтвердите допущения (если они есть), лежащие в основе критерия (например, нормальность, постоянство дисперсии), и опишите любые преобразования, требуемые для того, чтобы удовлетворить этим допущениям (например, взятие логарифмов).
- Определите наблюдаемую величину статистики критерия, его распределение (и степени свободы, если уместно) и, если возможно, точное значение достигнутого уровня значимости P (например, P=0,03), а не его интервальную оценку (например, $0{,}01<P<0{,}05$) или систему звёздочек (например, *, **, *** для увеличивающихся уровней значимости). Избегайте написания «н.з.» (не значимо), когда $P>0{,}05$; точное значение достигнутого уровня значимости P предпочтительнее, даже когда результат незначим.
- Включайте оценку относящегося к делу и интересующего вас эффекта (например, разность в средних для t-критерия двух выборок или средняя разность для парного t-критерия) с доверительным интервалом (предпочтительно) или стандартной ошибкой.
- Делайте заключения из результатов (например, отбросить нулевую гипотезу), интерпретируйте любой доверительный интервал и объясняйте их значения.

Результаты регрессионного анализа

Здесь мы учитываем парную и множественную линейную регрессию, логистическую регрессию, пуассоновскую регрессию, регрессию пропорциональных рисков и регрессионные методы для сгруппированных данных. Подробные детали этих анализов объяснены в соответствующих главах.

- Включайте уместные диаграммы (например, график рассеяния с предсказанной линией для простой регрессии).
- Отчетливо формулируйте, какая переменная зависимая, а какая — независимая.
- Подтверждайте допущения, лежащие в основе анализа.
- Описывайте любые преобразования и объясняйте их цель.
- Где уместно, описывайте возможные числовые значения, придаваемые любой категориальной переменной (например, мужчина=0, женщина=1), как были созданы (см. главу 29) искусственные переменные и единицы размерности непрерывных (количественных) переменных.
- Давайте показатель качества согласия модели (например, приведите R^2).
- Если уместно (например, во множественной регрессии), давайте результаты общего F-критерия из таблицы дисперсионного анализа (ANOVA).
- Давайте оценки всем коэффициентам в модели (включая незначимые) вместе с доверительными интервалами для коэффициентов или стандартные ошибки их оценок. При логистической регрессии (см. главу 30), пуассоновской регрессии и регрессии пропорциональных рисков преобразовывайте коэффициенты до оценок отношений шансов или относительных рисков (с доверительными интервалами). Интерпретируйте необходимые коэффициенты.
- Показывайте результаты проверки гипотез для коэффициентов (т.е. включайте статистики критерия и значения достигнутого уровня значимости P). Делайте соответствующие выводы из этих критериев.

Сложные анализы

Нет простых правил для представления более сложных форм статистического анализа. Убедитесь, что план исследования описан полностью (например, факторы в дисперсионном анализе, иерархическая классификация), и включайте подтверждение допущений, лежащих в основе необходимых описательных статистик (с доверительными интервалами), статистику критерия и значения достигнутого уровня значимости P. Краткое описание сути анализа помогает непосвящённым (читателям): оно должно сопровождаться ссылкой на будущие детали. Укажите, какая компьютерная программа использовалась.

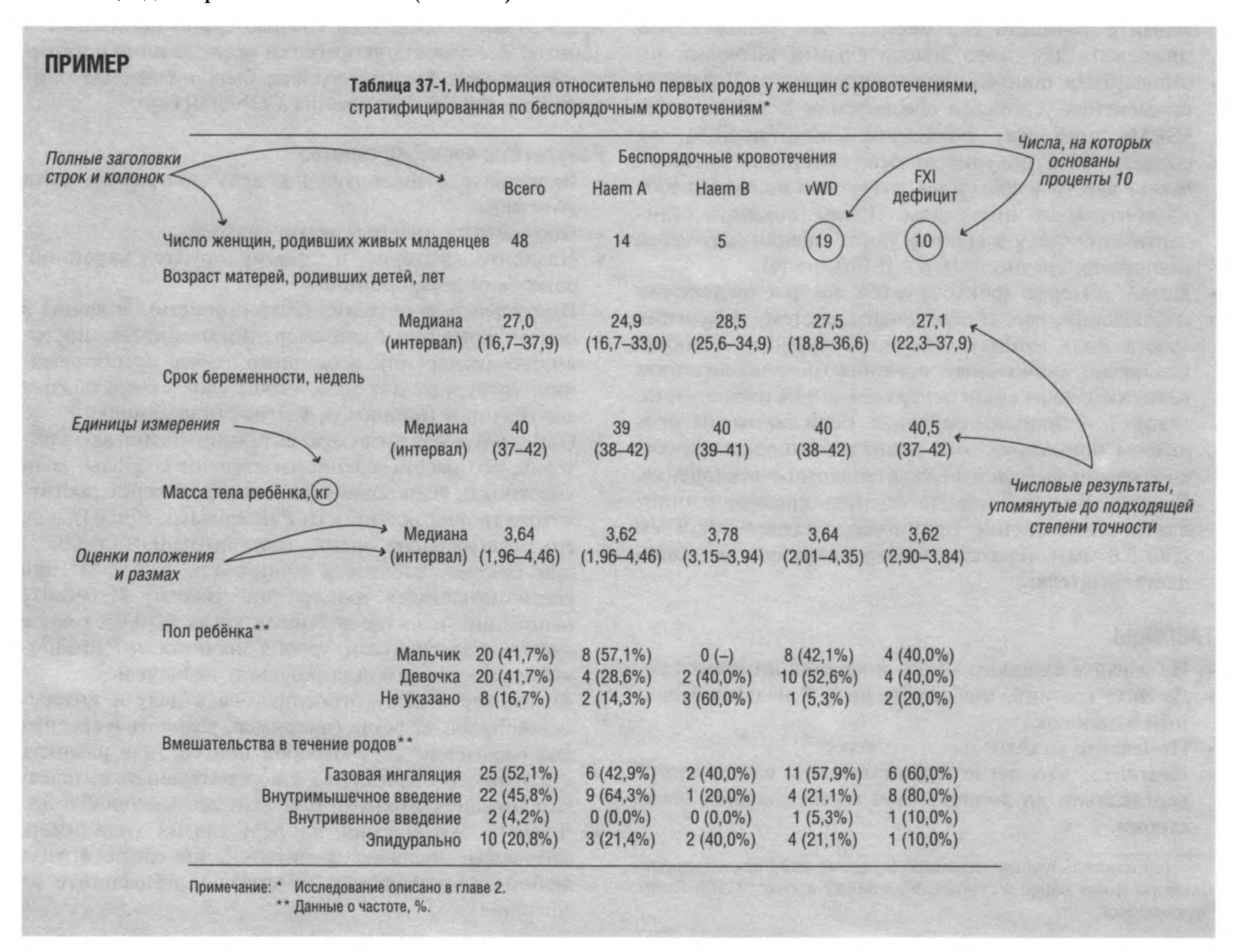


Таблица 37-1. Информация относительно первых родов у женщин с кровотечениями, стратифицированная по беспорядочным кровотечениям*

		Всего	Беспорядочные кровотечения Haem A	Haem B	vWD	FXI дефицит
Число женщин, родивших живых младенцев		48	14	5	19	10
Возраст матерей, родивших детей, лет						
	Медиана	27,0	24,9	28,5	27,5	27,1
	(интервал)	(16,7–37,9)	(16,7–33,0)	(25,6–34,9)	(18,8–36,6)	(22,3–37,9)
Срок беременности, недель						
	Медиана	40	39	40	40	40,5
	(интервал)	(37–42)	(38–42)	(39–41)	(38–42)	(37–42)
Масса тела ребёнка, кг						
	Медиана	3,64	3,62	3,78	3,64	3,62
	(интервал)	(1,96–4,46)	(1,96–4,46)	(3,15–3,94)	(2,01–4,35)	(2,90–3,84)
Пол ребёнка**						
	Мальчик	20 (41,7%)	8 (57,1%)	0 (–)	8 (42,1%)	4 (40,0%)
	Девочка	20 (41,7%)	4 (28,6%)	2 (40,0%)	10 (52,6%)	4 (40,0%)
	Не указано	8 (16,7%)	2 (14,3%)	3 (60,0%)	1 (5,3%)	2 (20,0%)
Вмешательства в течение родов**						
	Газовая ингаляция	25 (52,1%)	6 (42,9%)	2 (40,0%)	11 (57,9%)	6 (60,0%)
	Внутримышечное введение	22 (45,8%)	9 (64,3%)	1 (20,0%)	4 (21,1%)	8 (80,0%)
	Внутривенное введение	2 (4,2%)	0 (0,0%)	0 (0,0%)	1 (5,3%)	1 (10,0%)
	Эпидурально	10 (20,8%)	3 (21,4%)	2 (40,0%)	4 (21,1%)	1 (10,0%)

Примечание: * Исследование описано в главе 2.
** Данные о частоте, %.

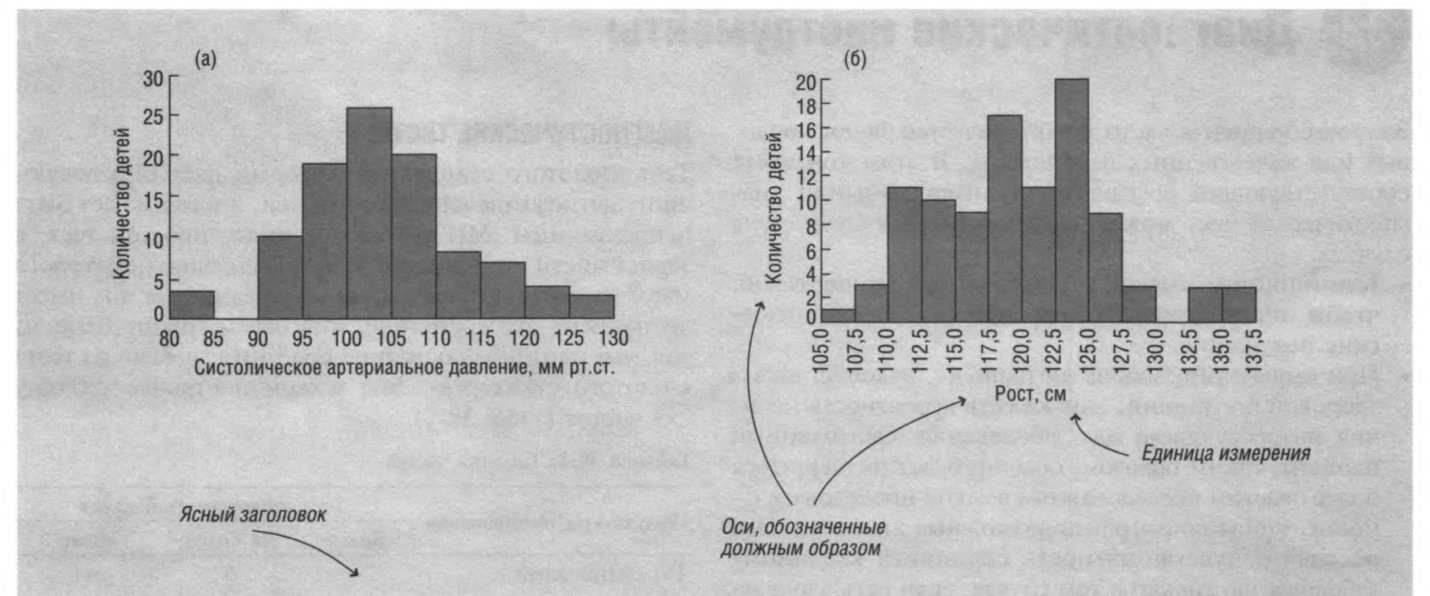


Рис. 37-1. Гистограммы, показывающие распределение: а) систолического артериального давления и б) рост в выборке из 100 детей.

38 Диагностические инструменты

Здоровье пациента часто характеризуется рядом числовых или качественных переменных. В этом контексте соответствующий референтный интервал и/или диагностический тест может использоваться в следующих случаях.

- Клиницистом вместе с клинической экспертизой, чтобы диагностировать или исключить специфические нарушения.
- При выяснении, здоров ли пациент, находящийся в здоровой популяции, или же есть вероятность наличия интересующего нас заболевания. Ослабленный пациент, таким образом, обычно будет подвергаться более строгим исследованиям в такой последовательности, чтобы подтвердить возможные диагнозы. Это обеспечит чувствительность скрининга к данному заболеванию только в том случае, если есть адекватные средства, возможности, которые помогут обеспечить лечение заболевания на стадии первичных симптомов. Именно это лечение менее дорогостоящее и/или более эффективно, чем лечение в более поздней стадии (или иногда, если предполагается, что пациенты, которые диагностированы как больные, изменят своё поведение, чтобы предотвратить развитие болезни).

РЕФЕРЕНТНЫЕ ИНТЕРВАЛЫ

Референтный интервал (часто называемый размахом, отвечающим нормальному распределению) для одной числовой переменной, рассчитанный из очень большой выборки, даёт диапазон величин, часто наблюдаемый у здоровых пациентов. Если индивидуальное значение выше верхнего предела такого интервала (или ниже нижнего), мы считаем его необычно высоким (или низким) относительно здоровых пациентов.

Расчёт референтных интервалов

Можно использовать два подхода.

- Мы допускаем, что данные нормально распределены. Около 95% значений данных лежит в пределе 1,96 стандартных отклонений среднего. Мы применяем наши данные для расчёта этих двух границ (среднее ±1,96×стандартное отклонение).
- Альтернативный подход, который не делает какихлибо допущений относительно распределения измерения, — применить центральный диапазон, который охватывает 95% величин данных. Мы располагаем наши значения в порядке величины и применяем 2,5 и 97,5 процентили в качестве наших пределов.

Влияние других факторов на референтные интервалы

Иногда значения числовой переменной зависят от других факторов, таких, как возраст или пол. Важно интерпретировать отдельное значение только после учёта этих других факторов. Например, мы генерируем референтные интервалы для САД отдельно для мужчин и женщин.

ДИАГНОСТИЧЕСКИЕ ТЕСТЫ

Тест «золотого стандарта», который даёт определяющий диагноз отдельного состояния, иногда может быть непрактичным. Мы хотели бы иметь простой тест, в зависимости от наличия или отсутствия некоторого маркёра, который даёт точное указание на то, имеет ли пациент это состояние. Мы берём группу пациентов, чьё истинное состояние болезни известно из теста «золотого стандарта». Мы можем построить таблицу 2×2 частот (табл. 38-1).

Таблица 38-1. Таблица частот

Результаты тестирования	Тест «золотого стандарта»		
	Болен	**Не болен**	**Итого**
Положительный	a	b	$a+b$
Отрицательный	c	d	$c+d$
Итого	$a+c$	$b+d$	$n=a+b+c+d$

Из исследованных n пациентов $a+c$ имеют заболевание. Распространённость (см. главу 12) заболевания в этой выборке $=(a+c)/n$. Из $(a+c)$ пациентов, которые имеют заболевание, а имеют положительные результаты тестирования (истинно положительные), а c имеют отрицательные результаты тестирования (ложноотрицательные). Из $b+d$ пациентов, которые не имеют болезни, d имеют отрицательные результаты тестирования (истинно отрицательные), а b — положительные результаты тестирования (ложноположительные).

Оценка надёжности: чувствительность и специфичность

Чувствительность = доля пациентов с заболеванием, которые точно идентифицированы тестом $=a/(a+c)$.

Специфичность = доля пациентов без заболевания, которые точно идентифицированы тестом $=d/(b+d)$.

Они обычно выражаются в виде процентов. Как и для всех оценок, мы должны рассчитать доверительные интервалы для этих измерений.

Мы хотели бы иметь чувствительность и специфичность, которые были бы обе как можно ближе к 1 (или 100%), если возможно. Однако на практике можно добиться чувствительности за счёт специфичности, и наоборот. Достигаем ли мы высокой чувствительности или высокой специфичности, зависит от состояния, которое мы пытаемся обнаружить, вместе с вовлечением для пациента и/или популяции ложноотрицательного или ложноположительного результата тестирования. Для состояний, которые легко излечиваются, мы предпочитаем высокую чувствительность; для состояний же, которые серьёзны и трудно излечимы, мы предпочитаем высокую специфичность, чтобы устранить возможную постановку ложноположительного диагноза. Важно, что прежде, чем начать скрининг, испытатели должны понять значения положительного диагноза, так же как наличие процента ложноположительных и ложноотрицательных результатов испытания.

Значения прогностичности

Прогностичность положительного результата теста = доля пациентов с положительным результатом диагностического теста, которые имеют заболевание $=a/(a+b)$.

Прогностичность отрицательного результата теста = = доля пациентов с отрицательным результатом критерия, которые не имеют заболевания $=d/(c+d)$.

Мы рассчитываем доверительные интервалы для этих прогностичных величин, часто выраженных как процентные соотношения, применяя методы, описанные в главе 11.

Эти прогностичные величины дают информацию о том, с какой вероятностью имеет или не имеет пациент заболевание при условии результата диагностического теста. Прогностичные величины зависят от распространённости болезни в исследуемой популяции. В популяциях, где болезнь встречается часто, положительная прогностичная величина будет гораздо выше, чем в популяциях, где болезнь редка. Обратное верно для прогностичных величин отрицательных результатов теста.

Использование «точек разделения»

Иногда мы хотим поставить диагноз на основе измерения непрерывных (количественных) величин. Часто не бывает порога, выше (или ниже) которого болезнь точно существует. В этих ситуациях нам нужно определить «точку разделения», выше (или ниже) которой, как мы полагаем, есть очень большая вероятность, что пациент болен.

Полезный подход — применить нижний (или верхний) предел референтного интервала. Можно оценить эту «точку разделения» путём расчёта соотносящихся с ней чувствительности, специфичности и прогностичности. Если мы выбираем различные точки деления, эти величины могут изменяться, по мере того как мы становимся более или менее строгими. Мы выбираем точку деления, исходя из желательной оптимизации этой системы мер.

Кривые операционной характеристики

Они дают способ оценки, обеспечивает ли данный тип диагностического теста полезную информацию и могут ли использоваться для сравнения двух различных тестов и для выбора значения оптимальной точки разделения для диагностического теста.

Для данного диагностического теста мы рассматриваем все точки разделения, которые дают уникальную пару значений для чувствительности и специфичности, и вычерчиваем зависимость чувствительности от величины 1 минус специфичность (сравнивая таким образом вероятности положительного результата диагностического теста у больных и здоровых пациентов) и соединяем эти точки линиями (рис. 38-1).

Кривая операционной характеристики диагностического теста (ROC) будет расположена слева от диагонали графика. Два или более теста можно сравнивать путём рассмотрения площади под каждой кривой — тест с большей площадью лучше[1]. В зависимости от значений ложноположительных и ложноотрицательных результатов теста и распространённости состояния из этого графика мы можем выбрать оптимальное значение точки разделения для теста. Два или более тестов (испытаний) на то же самое состояние можно сравнить, рассматривая область под каждой кривой: эта область даётся вместе со статистикой C (вычисляется многими статистическими пакетами). Испытание с большей областью (с большой величиной статистики C) лучше для различения между исходами болезни.

Критерий полезен?

Отношение правдоподобия (LR) для положительного результата теста есть отношение шанса положительного результата, если пациент болен, к шансу положительного результата, если он здоров (см. также главу 32). Отношения правдоподобия могут быть также генерированы для отрицательного результата теста. Например, LR, равное 2 для положительного результата, означает, что положительный результат, по-видимому, встречается у пациента с заболеванием в два раза чаще, чем у пациента без него. Высокое отношение правдоподобия для положительного результата предполагает, что этот тест даёт полезную информацию, равно как и отношение правдоподобия, близкое к нулю для отрицательного результата.

Можно показать, что:

LR для положительного результата = = (Чувствительность)/(1−специфичность).

Мы обсудим LR далее в главе 45.

[1] Поскольку площадь для каждой ROC является выборочной характеристикой, к тому же подчас полученным по выборкам разного объёма, то возможна проверка статистических гипотез о равенстве этих площадей для двух и более ROC. В некоторых современных статистических пакетах имеются такие процедуры. Прим. переводчика.

ПРИМЕР

Цитомегаловирус (CMV) — широко распространённая вирусная инфекция, которой около 50% пациентов подвержены в детстве. Хотя инфекция с этим вирусом часто не ведёт к серьёзным проблемам, пациенты, инфицированные CMV в прошлом, могут страдать от тяжёлых заболеваний после определённых процедур трансплантации, таких, как костная, если их вирус реактивируется или если они повторно будут инфицированы донорами. Было сделано предположение, что количество обнаруживаемого вируса после трансплантации (вирусная нагрузка) позволяет делать прогноз, какие субъекты тяжело заболеют. Чтобы исследовать эту гипотезу, вирусная нагрузка CMV была измерена в группе из 49 реципиентов костного трансплантата. У 15 из 49 развилось тяжёлое заболевание во время длительного наблюдения. Величины вирусной нагрузки варьировали от 2,7 lg genomes/мл до 6 lg genomes/мл. В качестве начальной точки рассматривалась величина более 4,5 lg genomes/мл как показатель возможного развития болезни в будущем. Нижеприведённая таблица частот показывает полученные результаты; график содержит расчёты интересующих характеристик.

Таблица 38-2. Результаты исследования прогноза возникновения заболевания после «вирусной нагрузки»

Вирусная нагрузка lg genomes/мл	Тяжёлое заболевание		
	Есть	Нет	Всего
>4,5	7	6	13
≤4,5	8	28	36
Всего	15	34	49

- Распространенность =15/49×100%=31% (95% ДИ от 18% до 45%).
- Чувствительность =7/15×100%=47% (95% ДИ от 22 до 72%).
- Специфичность =28/34×100%=82% (95% ДИ от 69 до 95%).
- Прогностичность положительного результата =7/13×100%=54% (95% ДИ от 27 до 81%).
- Прогностичность отрицательного результата =28/36×100%=78% (95% ДИ от 65 до 92%).
- Отношение правдоподобия для положительного результата =0,47/(1−0,82)=2,6 (95% ДИ от 1,1 до 6,5%, получено в результате компьютерного анализа).

Для этого значения точки разделения мы имеем относительно высокую специфичность и умеренную чувствительность. LR, равное 2,6, означает, что этот тест полезен тем, что вирусная нагрузка больше 4,5 lg genomes/мл более чем в 2 раза вероятна у пациента с тяжёлой болезнью, чем у пациента без неё. Однако, чтобы исследовать другие значения точки разделения, была построена кривая операционной характеристики (см. рис. 38-1). Нанесённая линия расположена левее диагонали графика. Для этого примера наиболее полезна точка разделения 5 lg genomes/мл, которая даёт чувствительность в 49% и специфичность в 97%; тогда LR равно 13,3.

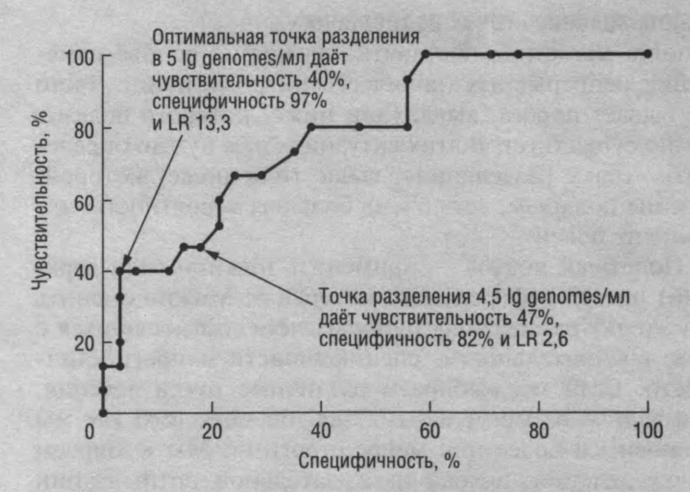


Рис. 38-1. Кривая операционной характеристики, показывающая результаты для двух возможных значений точек разделения, оптимальной и той, которая применялась в диагностическом тесте.

39 Оценка согласия

ВВЕДЕНИЕ

Есть много случаев, когда нужно сравнить результаты, которые должны совпадать. В частности, нужно оценить и, если возможно, количественно определить следующие два типа согласия.

- Воспроизводимость (согласие метод/наблюдатель). Создают ли два способа, применяемые для измерения отдельной переменной в идентичных условиях, те же результаты? Получают ли два или более наблюдателей одинаковые результаты, применяя один метод?
- Сходимость (повторяемость). Получает ли один наблюдатель те же результаты, когда он проводит повторные измерения в идентичных условиях?

Как воспроизводимость, так и сходимость можно достичь одним способом. В каждом случае метод анализа зависит от того, качественная (например, плохая/средняя/хорошая) переменная или числовая (например, САД). Для простоты мы ограничим проблему сравнением только двух групп парных результатов (т.е. два метода/два наблюдателя/двойные измерения).

КАЧЕСТВЕННЫЕ ПЕРЕМЕННЫЕ

Предположим, что два наблюдателя оценивают одних и тех же пациентов на предмет тяжести болезни, применяя качественную шкалу оценок, и мы хотим оценить степень, с которой эти оценки согласуются. Представляем результаты в двухмерной таблице сопряжённости частот с рядами и колонками, означающими категории ответа для каждого наблюдателя. Табл. 39-1 — пример, показывающий результаты работы двух наблюдателей относительно состояния поверхностей зубов. Частоты, когда оценки обоих наблюдателей идентичны, показаны вдоль диагонали таблицы. Рассчитываем соответствующие ожидаемые частоты, если бы классификации были установлены случайно таким же образом, как мы рассчитывали ожидаемые частоты в критерии Хи-квадрат Пирсона при анализе сопряжённости; т.е. каждая ожидаемая частота — произведение сумм частот соответствующих строк и колонок, делённого на общую сумму частот. Оцениваем согласие следующим образом (каппа Кохена):

$$\kappa = \frac{\frac{O_d}{m} - \frac{E_d}{m}}{1 - \frac{E_d}{m}}.$$

Выражение представляет собой шанс, скорректированный пропорционально согласию, где:

- m — общая наблюдаемая частота (т.е. общее число пациентов);
- O_d — сумма наблюдаемых частот вдоль диагонали;
- E_d — сумма ожидаемых частот вдоль диагонали;
- 1 в знаменателе представляет собой максимальное согласие.

$\kappa=1$ означает полное согласие, а $\kappa=0$ предполагает, что это согласие не лучше, чем согласие, которое может быть получено случайно. Для оценки промежуточных величин объективных критериев нет. Однако каппа часто считается обеспечивающей согласие[1], которое считают:

- плохим, если $\kappa \le 0{,}2$;
- допустимым, если $0{,}21 \le \kappa \le 0{,}4$;
- посредственным, если $0{,}41 \le \kappa \le 0{,}6$;
- значительным, если $0{,}61 \le \kappa \le 0{,}8$;
- хорошим, если $\kappa > 0{,}8$.

Хотя возможно оценить стандартную ошибку для каппы, мы обычно не проверяем гипотезу, что каппа равно нулю, так как это действительно не очень подходящий или реальный инструмент в изучении надёжности.

Отметьте, что каппа зависит как от числа категорий (т.е. её величина больше, если категорий меньше), так и распространённости состояния, поэтому следует быть осторожными при сравнении каппы из различных исследований. Для порядковых данных мы можем также рассчитать взвешенную каппу[2], которая принимает во внимание степень, с которой наблюдатели не согласны (недиагональные частоты), а также частоты согласия (на диагонали). Взвешенная каппа подобна внутриклассовому коэффициенту корреляции (см. следующий раздел и главу 42).

ЧИСЛОВЫЕ ПЕРЕМЕННЫЕ

Предположим, что наблюдатель берёт спаренные измерения числовой переменной от n пациентов (замените слово «повторяемость» на «воспроизводимость» при рассмотрении подобной проблемы согласия метода).

- Если средняя разность (т.е. истинная средняя разность, оценённая как $\bar{d}$) равна нулю (по данным парного t-критерия, знакового критерия или критерия знаковых рядов, главы 19 и 20), мы можем заключить, что в результатах нет смещения. Это подразумевает, что в среднем двойные измерения соответствуют друг другу.
- Оценка стандартного отклонения парных разностей (s_d) обеспечивает измерение согласия, которое можно использовать как сравнительный инструмент. Однако чаще рассчитывают коэффициент воспроизводимости Британского института стандартов $=2s_d$. Он показывает максимальную разность, которая, возможно, будет иметь место между двумя измерениями, если нет смещения. Предполагая нормальное распределение разности, мы ожидаем, что приблизительно 95% разницы в популяции лежит между $\bar{d} \pm 2s_d$. Верхний и нижний пределы этого интервала называются пределами согласия; из них мы можем решить (субъективно), приемлемо ли согласие между парами считываний в данной ситуации.

---

[1] Landis J.R., Koch G.G. The measurement of observer agreement for categorical data // Biometrics, 1977. — Vol. 33. — P. 159–174.

[2] Cohen J. Weighted kappa: nominal scale agreement withprovision for scale disagreement or partial credit // Psychological Bulletin, 1968. — Vol. 70. — P. 213–220.

- Индекс надёжности обычно используют для оценки воспроизводимости и повторяемости внутриклассового коэффициента корреляции (ВКК), который принимает значения от нуля (полное отсутствие согласия) до 1 (полное согласие). При оценке согласия между парами наблюдений ВКК характеризует долю изменчивости в наблюдениях, которая определяется различием между парами, т.е. это величина межпарного различия, выраженная как пропорция общей вариабельности всех наблюдений.

Когда нет никакого свидетельства систематического различия между парами, можно вычислять ВКК как коэффициент корреляции Пирсона между $2n$ пар наблюдений, полученных включением каждой пары дважды: первый раз, когда её значения наблюдаются в начале исследования, и повторно — при изменении условий (см. пример 2).

Если мы принимаем во внимание систематическое различие между парами наблюдений, мы оцениваем ВКК как:

$$\frac{s_a^2 - s_d^2}{s_a^2 + s_d^2 + \frac{2}{n}\left(n\bar{d}^2 - s_d^2\right)}.$$

При этом мы определяем различия и сумму наблюдений в каждой из n пары:

- s_a^2— оценка дисперсии n сумм;
- s_d^2 — оценка дисперсии n разностей (оценка систематической разности, смещения).

Мы обычно выполняем исследование надёжности как часть большого исследовательского этапа. Выборка наблюдений, используемая для изучения надёжности, должна соответствовать этой исследовательской стадии. Нельзя сравнить значения ВКК в различных наборах данных, поскольку ВКК находится под влиянием особенностей данных, типа их изменчивости (ВКК будет больший, если наблюдения имеют большее количество переменных). Кроме того, ВКК не связан с фактической шкалой измерения или величиной клинически приемлемой ошибки.

Предосторожности

- Нет смысла рассчитывать одну меру воспроизводимости, если степень, до которой не согласуются наблюдения в паре, зависит от величины измерения. Мы можем проверить это путём определения как средней, так и разности между каждой парой считываний и нанесения n разностей против их соответствующих средних[1] (рис. 39-1). Если соотношения нет, мы должны увидеть случайный разброс точек (распределённых равным образом выше и ниже нуля, если смещения нет). Однако если мы наблюдаем эффект воронки (колокола) с изменением разницы, которая для больших средних величин, скажем, больше, нужно повторно оценить проблему. Мы можем найти подходящее преобразование исходных данных так, что при повторении процесса на преобразованных наблюдениях требуемое условие будет удовлетворено. Для обнаружения аномальных значений мы можем также применить график.
- Будьте особенно внимательны при создании двухмерного графика с результатами от первого измерения, нанесёнными в зависимости от результатов второго измерения (или данных от одного метода/наблюдателя, нанесённых против другого), и рассчитывая коэффициент корреляции (см. главу 26). Практически мы не интересуемся, лежат ли эти точки на прямой линии; мы хотим знать, согласуются ли они с 45-градусной линией, т.е. линией равенства. То, что истинный коэффициент корреляции равен нулю, не будет установлено критерием проверки нулевой гипотезы. Кроме того, примите во внимание тот факт, что можно увеличить коэффициент корреляции путём увеличения диапазона величин измерений.

Более сложные ситуации

Иногда вы можете столкнуться с более сложными проблемами при оценке согласия. Например, может быть больше чем две повторности, или больше чем два наблюдателя, или кто-то из числа наблюдателей может иметь повторные наблюдения. Вы можете найти детали анализа таких проблем в Streiner and Norman[2].

ПРИМЕР 1

Оценка согласия — качественная переменная

Два наблюдателя, опытный дантист и студент стоматологического факультета, оценили состояние 2104 зубных поверхностей у детей школьного возраста. Каждая поверхность была закодирована как «0» (крепкие), «1» (по меньшей мере, с одной небольшой полостью), «2» (по меньшей мере, с одной большой полостью) или «3» (по меньшей мере, с одной пломбой, с полостью или без неё) у каждого пациента. Наблюдаемые частоты показаны в таблице 39-1. Выделенные цифры по диагонали показывают наблюдаемые частоты согласия; соответствующие ожидаемые частоты даны в скобках. Мы рассчитали каппу Кохена для оценки согласия между этими двумя наблюдателями:

$$\kappa = \frac{\left(\frac{1785+154+20+14}{2104}\right) - \left(\frac{1602{,}1+21{,}3+0{,}5+0{,}2}{2104}\right)}{1-\left(\frac{1602{,}1+21{,}3+0{,}5+0{,}2}{2104}\right)} = \frac{0{,}9377-0{,}7719}{1-0{,}7719} = 0{,}73.$$

Между студентом и опытным дантистом наблюдается большое согласие относительно кодирования состояния зубных поверхностей детей.

[1] Bland J.M., Altman D.G. Statistical methods for assessingagreement between two pairs of clinical measurement. Lancet, i. — P. 307–310.

[2] Streiner D.R., Norman G.L. Health measurement scales: A practical guide to their development and use. — 3 edn. — Oxford: Oxford University Press, 2003.

Таблица 39-1. Наблюдаемые (и ожидаемые) частоты оценки поверхности зуба

	Коды	Оценки студента				
		0	1	2	3	Итого
Оценки дантиста	0	**1785** (1602,1)	46	0	7	1838
	1	46	**154** (21,3)	18	5	223
	2	0	0	**20** (0,5)	0	25
	3	3	1	0	**14** (0,2)	18
	Итого	1834	201	43	26	2104

Данные любезно предоставлены Dr R.D. Holt. Eastman Dental Institute, University College London, London, UK.

ПРИМЕР 2

Оценка согласия — числовая переменная

Индекс самоуважения Розенберга применяют для того, чтобы судить об оценке пациента по его самооценке. Минимальное значение этого индекса (показатель высокой самооценки) для субъекта — 50, что составляет сумму индивидуальных оценок по 10 вопросам, по каждому из которых дан балл от 0 до 5. В части исследования эффективности операции по поводу деформации лица изучались изменения психологического профиля пациента путём сравнения величин индекса Розенберга у пациента до и после операции. Исследователи беспокоились о степени, до которой индекс Розенберга должен быть надёжен для группы пациентов, и решили оценить повторяемость измерения на первых 25 пациентах, требующих лечения по поводу деформации лица. Они получили величину индекса Розенберга, когда пациент поступил в клинику, а затем попросили у пациента вторую оценку через четыре недели. Результаты показаны в табл. 39-2.

Можно показать, что разность (первое значение минус второе значение) приблизительно нормально распределена; она имеет среднее $\bar{d}$=0,56 и стандартное отклонение s_d=1,83. Статистика критерия для парного t-критерия равна 1,53 (степени свободы =24), давая P=0,14. Этот результат означает, что данных о каком-либо смещении нет. Британский стандарт коэффициента повторяемости, равный $2 \times s_d = 2 \times 1{,}83 = 3{,}7$. Следует ожидать, что около 95% разности в популяции таких пациентов будет лежать между $\bar{d} \pm 2s_d$, т.е. между −3,1 и 4,3. Эти пределы даны на рис. 39-1, который показывает, что разность имеет случайный разброс вокруг среднего, равного приблизительно нулю. На основе этих результатов исследователи почувствовали, что индекс Розенберга надёжен, и применили его, чтобы оценить восприятие пациентами эффективности косметической хирургии.

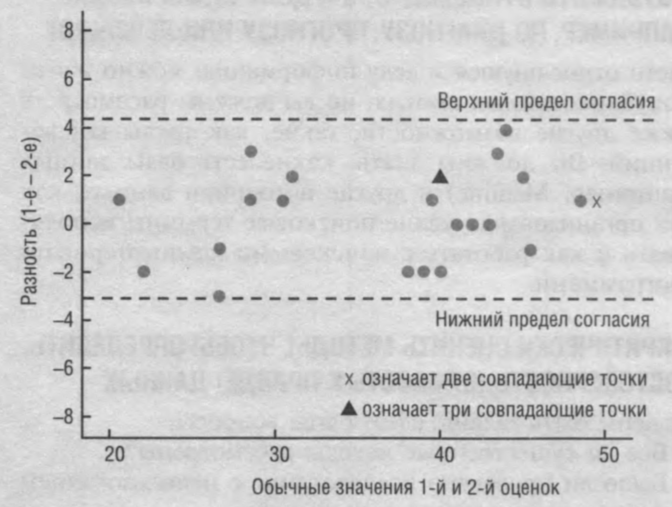


Рис. 39-1. Разность между первой и второй величинами индекса самооценки Розенберга в зависимости от их средней для 25 пациентов.

Таблица 39-2. Долечебные оценки (1-е и 2-е) индекса Розенберга, полученные от 25 пациентов

1-я оценка	2-я оценка	1-я оценка	2-я оценка	1-я оценка	2-я оценка	1-я оценка	2-я оценка	1-я оценка	2-я оценка
30	27	41	39	37	39	43	43	21	20
39	41	41	41	42	42	40	39	41	39
50	49	50	49	46	44	31	30	29	28
45	42	38	40	49	48	45	46	26	27
25	28	41	39	21	23	46	42	32	30

Адаптировано из: Cunningham S.J., Hunt N.P., Feinnman C. Perceptions of outcome following orthognathic surgery // British Journal of Oral ami Maxillofacial Surgery, 1996. — Vol. 34. — P. 210−213.

40 Доказательная медицина

Sackett et al[1] описывают доказательную медицину (ДМ) как «добросовестное, точное и здравомыслящее применение лучших современных доказательств при принятии решения о лечении отдельных пациентов». Чтобы применять ДМ на практике, вы должны уметь разработать исследование, подходящее для лечения ваших пациентов, и оценить его качество. Только тогда вы можете думать о применении находок в клинической практике.

Sackett et al предлагают следующий подход к ДМ. Для удобства мы сформулировали третий и четвёртый пункты в рамках клинических и наблюдательных исследований, но их можно модифицировать, чтобы они подходили к другим видам исследований (например, диагностические критерии).

1. СФОРМУЛИРОВАТЬ ПРОБЛЕМУ

Вы должны решить, что вас интересует — как вы определяете популяцию пациентов, какое вмешательство (т.е. лечение) или сравнение уместно и какой исход вы ожидаете (например, снижение смертности).

2. ОТЫСКАТЬ ОТНОСЯЩУЮСЯ К ДЕЛУ ИНФОРМАЦИЮ (НАПРИМЕР, ПО ДИАГНОЗУ, ПРОГНОЗУ ИЛИ ЛЕЧЕНИЮ)

Часто относящуюся к делу информацию можно найти в опубликованных статьях, но вы должны рассмотреть также другие возможности, такие, как труды конференций. Вы должны знать, какие есть базы данных (например, Medline) и другие источники данных, как они организованы, какие поисковые термины использовать и как работать с поисковыми компьютерными программами.

3. КРИТИЧЕСКИ ОЦЕНИТЬ МЕТОДЫ, ЧТОБЫ ОПРЕДЕЛИТЬ ДОСТОВЕРНОСТЬ (БЛИЗОСТЬ К ПРАВДЕ) ДАННЫХ

Должны быть заданы следующие вопросы.

- Все ли существенные исходы рассмотрены?
- Было ли проведено исследование с использованием подходящего спектра пациентов?
- Имеют ли результаты биологический смысл?
- Было ли исследование разработано так, чтобы устранить смещение? Например, в клиническом исследовании: было ли исследование контролируемым, использовалась ли рандомизация при отборе пациентов, была ли оценка ответу «слепой», были ли потеряны пациенты во время длительного наблюдения, в равном ли положении находились группы пациентов, помимо того что они получали различное лечение, и был ли проведён анализ фактически полученного ими лечения?
- Подходят ли статистические методы (например, подтверждены ли лежащие в их основе допущения; приняты ли во внимание в анализе зависимости данных, например, образование пар наблюдений)?

4. ИЗВЛЕЧЬ НАИБОЛЕЕ ПОЛЕЗНЫЕ РЕЗУЛЬТАТЫ И ОПРЕДЕЛИТЬ, ЗНАЧИТЕЛЬНЫ ЛИ ОНИ

Извлечение наиболее полезных результатов. Вы должны задать следующие вопросы:

- 1. Что является основной переменной результата (т.е. что относится к основной цели)?
- 2. Насколько велик интересующий нас эффект, выраженный в терминах основной переменной результата?
  - ✧ Если эта переменная бинарная (например, умер/живой).
    - ✦ Какова частота этого явления (например, смерти) в (двух) группах сравнения?
    - ✦ Каковы размеры интересующего нас эффекта, который может быть разностью в частотах (абсолютное уменьшение риска) или соотношением частот (относительный риск или отношение шансов)?
  - ✧ Если эта переменная числовая (например, САД).
    - ✦ Какова средняя величина переменной в каждой из групп сравнения?
    - ✦ Каков интересующий нас эффект, например, разность в средних?
- 3. Насколько точен интересующий нас эффект? В идеале исследование при внимательном его рассмотрении должно содержать доверительный интервал для истинного эффекта (широкий доверительный интервал — показатель плохой точности). Делается ли ссылка на этот доверительный интервал? Если нет, обеспечен ли достаточный объём информации (например, стандартная ошибка интересующего нас эффекта) так, чтобы доверительный интервал был определён?

Решение о том, значимы ли результаты

- Рассмотрите доверительный интервал для интересующего нас эффекта (например, разности в средних показателях лечения):
  - ✧ Считали бы вы наблюдаемый эффект клинически значимым (независимо от того, значим или нет статистически уместный критерий проверки гипотез), если бы нижний предел доверительного интервала представлял собой истинную величину этого эффекта?
  - ✧ Считали бы вы наблюдаемый эффект клинически значимым, если бы верхний предел доверительного интервала представлял собой истинную величину этого эффекта?
  - ✧ Достаточно ли одинаковы ваши ответы на первые два вопроса, чтобы объявить результаты исследования недвусмысленными и важными?
- Чтобы оценить лечение в рандомизированном контролируемом исследовании, определите число больных, которое вам необходимо лечить (ЧБНЛ) при

[1] Sackett D.L., Richardson W.S., Rosenberg W., Haynes R.B. Evidence–based Medicine: How to Practice and Teach EBM. London: Churchill–Livingstone, 1997.

экспериментальном, а не контролируемом лечении, чтобы не допустить у одного из них развития «плохого» исхода (такого, как послеродовое кровотечение, см. пример). ЧБНЛ можно определить различными способами в зависимости от имеющейся информации. Это, например, обратная величина разности в долях пациентов с плохим исходом в контроле и в экспериментальных группах (см. пример).

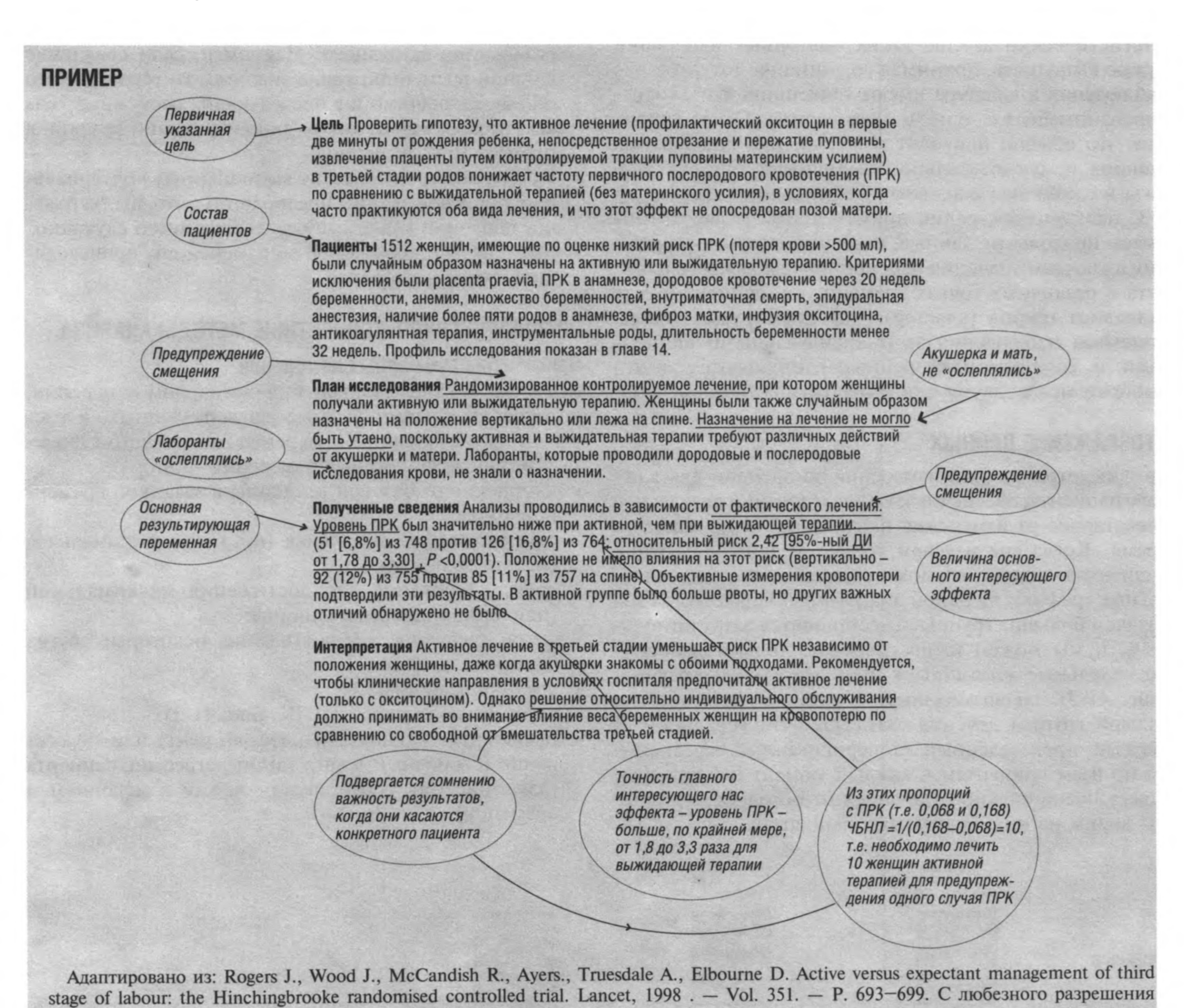
ПРИМЕНЕНИЕ РЕЗУЛЬТАТОВ В КЛИНИЧЕСКОЙ ПРАКТИКЕ

Если результаты должны помочь вам в лечении ваших пациентов, вы должны гарантировать, что:

- ваш пациент подобен пациентам, на которых были получены эти результаты;
- результаты могут быть применены к вашему пациенту;
- рассмотрены все клинически значимые исходы;
- предположительная выгода стоит потенциального ущерба и затрат.

ОЦЕНИТЬ СВОЁ ИСПОЛНЕНИЕ

Самооценка заключается в том, чтобы проверить свои возможности в успешном выполнении пунктов 1–5. Способны ли вы затем интегрировать критическую оценку в клиническую практику и проверили ли вы свои способности? Вы должны также спросить себя, извлекли ли вы что-нибудь из прошлого опыта, более ли эффективны вы сейчас и находите ли весь процесс ДМ более лёгким.



Адаптировано из: Rogers J., Wood J., McCandish R., Ayers., Truesdale A., Elbourne D. Active versus expectant management of third stage of labour: the Hinchingbrooke randomised controlled trial. Lancet, 1998 . — Vol. 351. — P. 693–699. С любезного разрешения издательства Elsevier.

41 Методы для сгруппированных данных

Сгруппированные данные соответствуют иерархической или вложенной структуре, в которой в её самой простой форме (одномерная структура с двумя уровнями) значения единственной переменной отклика (зависимой переменной) измерены на первом уровне группирующей переменной, в отличающиеся от второй группы или кластера. Например, уровнем 1-й и уровнем 2-й единицы измерения соответственно могут быть зубы во рту, коленный сустав у пациента, пациенты в больнице, клиники в регионе, дети в классе, последовательные времена посещения пациентом лечебного учреждения (т.е. продольные данные, рис. 41-1) и т.д. Статистический анализ таких повторных измерений нужно выполнять, принимая во внимание тот факт, что наблюдения в кластере имеют тенденцию быть коррелированными, т.е. они не независимы. Отказ учитывать это обычно приводит к недооценке стандартных ошибок и, следовательно, увеличивает ошибку 1-го рода и слишком узкие доверительные интервалы.

С целью иллюстрации примем в этой главе, что мы имеем продольные данные, и наши повторные измерения включают значения переменной для каждого пациента в различных точках времени, т.е. пациент и есть отдельная группа (кластер). Мы суммируем данные, описывая наблюдения за индивидуальными пациентами, и, если уместно, оцениваем, отличаются ли эти выборки между двумя или более группами пациентов.

ОТОБРАЖЕНИЕ ДАННЫХ

График, отражающий измерения во времени для каждого пациента, обеспечивает в исследовании визуальное впечатление от изменения наблюдений через какое-то время. Когда мы изучаем только маленькую группу пациентов, вполне возможно показать все индивидуальные графики на одной диаграмме. Однако когда мы изучаем большие группы, это становится затруднительным, и мы можем иллюстрировать только выборочно отдельные «типичные» индивидуальные графики (рис. 41-3), по возможности в едином масштабе для каждой группы лечения. Заметьте, что усреднённый образец, произведённый, сгенерированный усреднением по всем пациентам в каждый момент наблюдения, может очень сильно отличаться от образцов, которые мы видим на индивидуальных пациентах.

СРАВНЕНИЯ ГРУПП: НЕПОДХОДЯЩИЕ МЕТОДЫ АНАЛИЗА

Некорректно приспособить к таким данным парную линейную регрессию или проводить однофакторный дисперсионный анализ (ANOVA), применяя все величины, поскольку эти методы не принимают во внимание повторные измерения у одного и того же пациента. Более того, также некорректно сравнивать средние в группах по каждой временной точке отдельно, применяя непарный *t*-критерий или однофакторный ANOVA, по ряду причин.

- Измерения у пациента от одной временной точки до другой не независимы, так что интерпретация результатов затруднена. Например, если сравнение в одной временной точке значимо, то вероятно, что оно будет значимо и в последующих временных точках, независимо от каких-либо изменений величин в промежуточный период.
- По большому количеству выполненных критериальных проверок можно предполагать, что мы, вероятно, получили значимые результаты чисто случайно.
- Мы теряем информацию об изменениях, происходящих в самом пациенте.

СРАВНЕНИЯ ГРУПП: КОРРЕКТНЫЕ МЕТОДЫ АНАЛИЗА

Применение суммарных измерений

Мы можем основать анализ на суммарном измерении, которое захватывает важные аспекты данных, и рассчитать это суммарное измерение для каждого субъекта. Типичные суммарные измерения — это:

- изменение от базовой величины в заданной временной точке;
- достигнутая максимальная (пик) или минимальная (надир) величина;
- время, требуемое для достижения максимальной (или минимальной) величины;
- время, требуемое для достижения некоторых других заданных величин;
- средняя величина;
- площадь под кривой (ППК, рис. 41-2);
- наклон (регрессионный коэффициент) или пересечение (свободный член) линии регрессии пациента (для описания соотношения между измерением и временем).

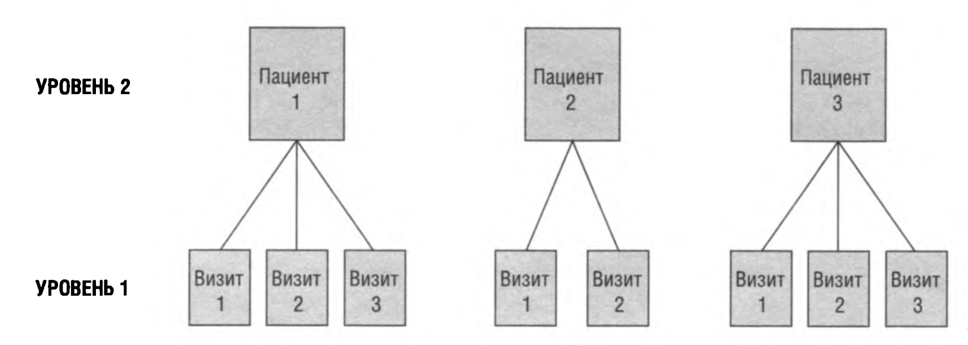


Рис. 41-1. Схематическое представление иерархической структуры с двумя уровнями для продольных данных.

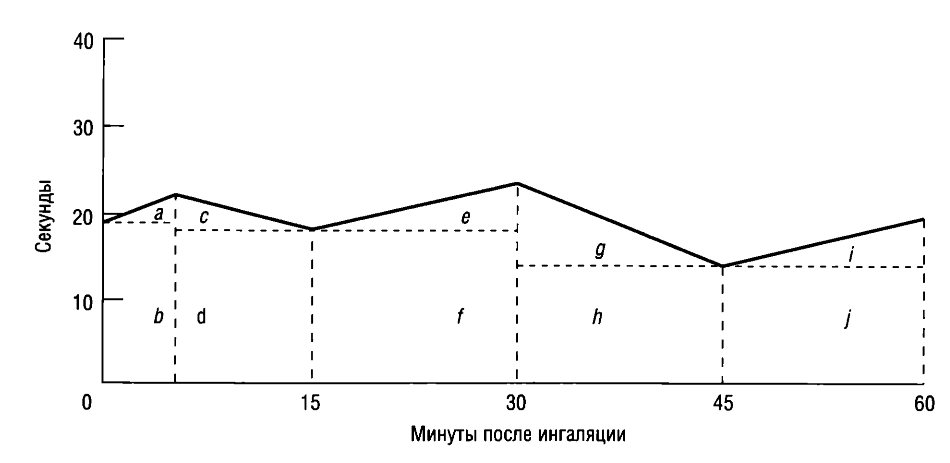


Рис. 41-2. Расчёт площади под кривой для одного студента. Общую площадь под линией можно разделить на ряд четырёхугольников и треугольников (обозначенных от *a* до *j*). Площадь каждого можно легко рассчитать. Общая площадь под кривой = площадь (*a*)+площадь (*b*)+...+площадь (j).

Выбор суммарного измерения зависит от основного интересующего нас вопроса и должен быть сделан до сбора данных. Например, если мы рассматриваем концентрации препаратов после лечения двумя способами, можно выбрать время, прошедшее до максимальной концентрации препарата (C_{max}) или ППК. Однако если нас интересуют титры антител после вакцинации, тогда мы можем выбрать время, требуемое титру антитела, чтобы опуститься ниже определённого защитного уровня.

Мы сравниваем эти величины суммарного измерения в различных группах, применяя стандартные критерии проверки гипотез (например, двухвыборочный критерий Вилкоксона ранговых сумм или критерий Крускала–Уоллиса). Поскольку мы уменьшили число зависимых измерений по каждому субъекту до одного значения, то величины, включённые в анализ, теперь независимы.

Хотя анализы, основанные на суммарных измерениях, проводить легко, довольно трудно найти подходящее измерение, которое адекватно описывает данные, и нам может потребоваться применить два или более суммарных измерения. Кроме того, эти подходы не используют полностью все значения данных.

Дисперсионный анализ повторных измерений

Мы можем провести особый тип ANOVA, называемый ANOVA повторных измерений, при котором различные временные точки считаются уровнями одного фактора в анализе, а групповая переменная — вторым фактором в анализе. Если этот анализ создаёт значимые разницы между группами, тогда могут быть использованы скорректированные *t*-критерии, которые принимают во внимание зависимость данных, чтобы идентифицировать, в каких временных точках эти различия становятся несомненными[1]. Однако у ANOVA повторных измерений есть некоторые недостатки.

- Часто его трудно выполнять.
- Результаты бывает трудно интерпретировать.
- Он в целом допускает, что величины измеряют с регулярными (эквидистантными) временны́ми интервалами и что нет потерянных данных, т.е. допускается, что план исследования сбалансирован. В действительности же величины редко измеряют во всех равноудалённых временных точках, поскольку пациенты часто пропускают назначенные встречи или приходят не в то время, когда планировалось.

Регрессионные методы

Различные методы регрессионного анализа, подобные тем, которые обеспечивают оценки параметра устойчивыми стандартными ошибками или используют ППК или случайные модели эффектов, можно использовать, чтобы проанализировать сгруппированные данные.

[1] Mickey R.M., Dunn O.J., Clark V.A. Applied Statistics: Analysis of Variance and Regression — 3 edn. — Wiley, 2004.

ПРИМЕР

Как часть практического курса обучения, разработанного для оценки эффектов двух вдыхаемых бронходилататоров, фенотерол гидробромида и ипратропиум бромида, 99 студентов-медиков были рандомизированы на получение одного из этих препаратов (n=33 для каждого препарата) или плацебо (n=33). Каждый студент вдыхал их четыре раза в быстрой последовательности. Тремор оценивался путём измерения общего времени (в секундах), необходимого для продевания нитки в пять швейных игл, закреплённых на пробке; измерения делались до ингаляции и через 5, 15, 30, 45 и 60 мин после неё. Измерения репрезентативной выборки студентов в каждой группе лечения показаны на рис. 41-2.

Было решено сравнить величины в этих трёх группах, используя ППК как суммарное измерение. Расчет ППК для одного студента показан на рис. 41-3.

Медиана (размах) ППК были соответственно 1552,5 (417,5–3875), 1215 (457,5–2500) и 1130 (547,5–2625) c^2 у студентов, получающих эти два препарата и плацебо. Величины в этих трёх группах сравнивались с применением критерия Краскел–Уоллиса, который дал P=0,008. Имелись, таким образом, убедительные аргументы, что измерения ППК в трёх группах были различными. Непараметрические апостериорные сравнения показали, что величины были больше в группе, получающей фенотерол гидробромид, подтверждая фармакологические сведения, что это лекарство как агонист β_2-адренорецептора вызывает тремор путём стимуляции β_2-адренорецепторов в скелетной мышце.

Данные были любезно предоставлены Dr R. Morris, Department of Primary Care and Population Sciences, and were collected as part of a student practical class organized by Dr T.J. Alien, Department of Pharmacology, Royal Free and University College Medical School, London, UK.

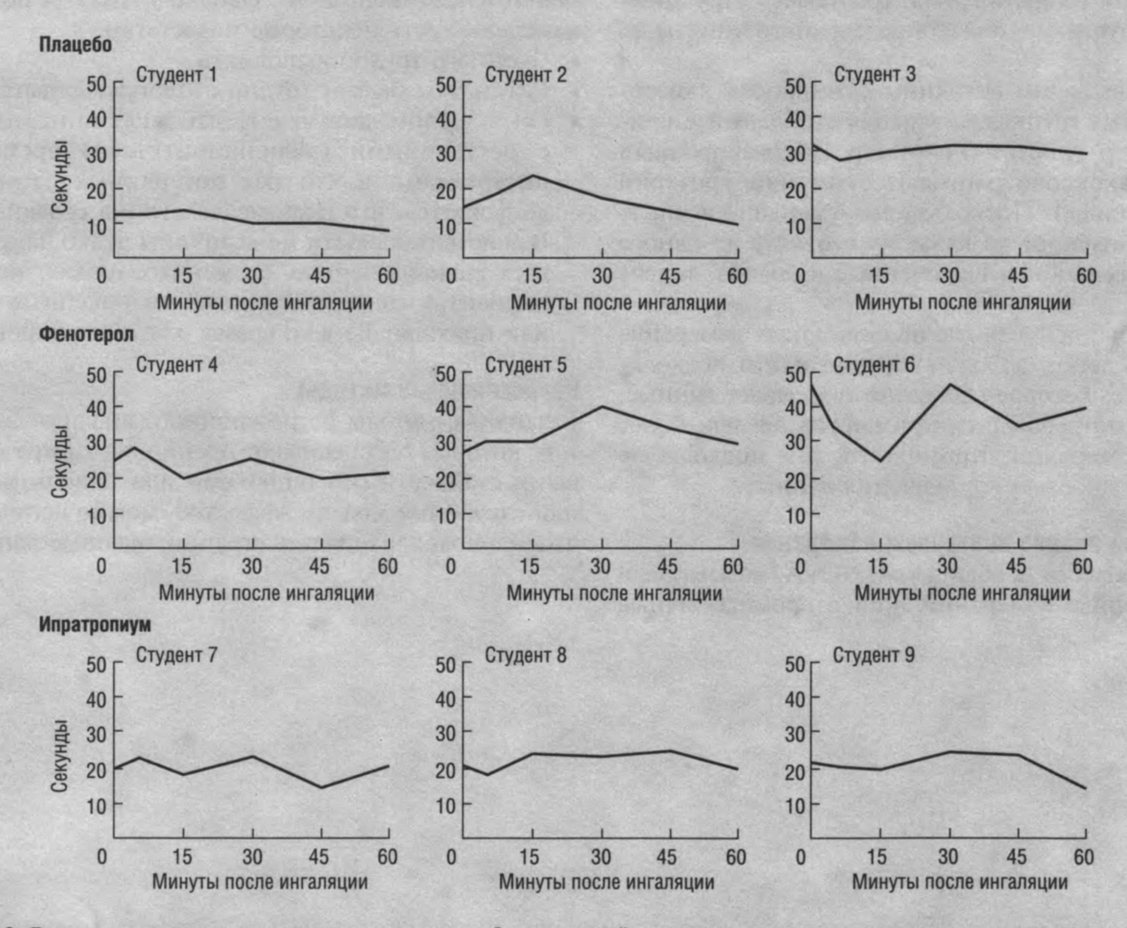


Рис. 41-3. Время, затраченное на продевание нитки в 5 игл для трёх репрезентативных студентов в каждой группе лечения.

42 Регрессионные методы для сгруппированных данных

Различные методы регрессионного анализа могут использоваться для анализа иерархической структуры с двумя уровнями, описанной в главе 41, в которой каждая группа (шкала второго уровня) содержит множество индивидуальных значений первого уровня. Например, при изучении ревматоидного артрита мы можем измерять угол сгибания на левых и правых коленях (шкала первого уровня) каждого пациента (шкала второго уровня). Или можно иметь продольный набор данных с измерением (например, общего холестерина), наблюдаемый в последовательные точки времени (шкала первого уровня) на каждом пациенте (шкала второго уровня). Главные преимущества и неудобства каждого метода обобщены в табл. 42-1. Большинство этих методов ненадёжно, если отсутствуют последовательные группы, и они могут быть усложнены, чтобы быть исполненными и правильно интерпретированными; поэтому лучше получить рекомендации специалиста-статистика.

АНАЛИЗ АГРЕГИРОВАННЫХ УРОВНЕЙ

Очень простой подход состоит в том, чтобы агрегировать данные и выполнить анализ, используя соответствующую числовую итоговую меру (например, среднее) для каждой группы (например, пациента). Выбор этой итоговой меры будет зависеть от особенностей данных и от изучаемых гипотез. Мы используем обычно МНК, многократно повторяя регрессионный анализ, используя группу как единицу наблюдения в исследовании и итоговую меру как результирующую переменную. Если каждая группа была размещена по специфическому виду лечения, (например, при лечении колена пациент может быть рандомизированно размещён по одному из двух режимов лечения или в группу без лечения, тогда вместе с другой группой, выравнивая по другому ковариату (например, пол, возраст), мы можем включать вид лечения в регрессионную модель как фиктивную переменную, используя коды типа 0 и 1 (или как ряд фиктивных переменных, если видов мы лечения более двух).

УСТОЙЧИВЫЕ СТАНДАРТНЫЕ ОШИБКИ

Если факт группировки в виде структуры с двумя уровнями игнорируется в регрессионном анализе, то важнейшее предположение, лежащее в основе линейной регрессионной модели, — наличие независимости между отдельными наблюдениями, нарушено. Как следствие этого нарушения стандартные ошибки оценок параметров, вероятно, будут слишком малы и, следовательно, результаты могут быть ложнозначимы.

Чтобы преодолеть эту проблему, можно определять устойчивые стандартные ошибки оценок параметров, базируя вычисление их на вариабельности данных (оценивая соответствующие остатки) точнее, чем по принятой регрессионной модели. Во множественном регрессионном анализе с устойчивыми стандартными ошибками оценки коэффициентов регрессии те же самые, что и в МНК линейном регрессионном анализе, но стандартные ошибки более устойчивы к нарушениям основных предположений, проявляющихся в недостаточной независимости, когда группируют данные.

МОДЕЛИ СЛУЧАЙНЫХ ЭФФЕКТОВ

Модели случайных эффектов[1], известные так же, как иерархические, многоуровневые, смешанные, специфически групповые или взаимно секционные модели временных рядов. Они могут быть реализованы с помощью различных комплексных статистических компьютерных пакетов типа SAS и Stata или специального программного обеспечения типа MLwiN (http://multilcvel.ioe.ac.uk), каждый из которых использует оценки максимального правдоподобия. Оценка эффекта для каждой группы получается благодаря использованию как индивидуальной информации по отдельной группе, так и информации по другим группам таким образом, чтобы это позволяло извлечь выгоду из «разделённой» информации. В частности, оценки потерь обычно определяют посредством того, что, используя соответствующий показатель потерь, оценивают долю вклада каждой группы в общую оценку. Величина потерь зависит от объёма наблюдений в каждой группе (группы с меньшими объёмами имеют большие потери) и вариабельности в данных (потери больше для оценок, когда вариация в пределах группы больше, чем между этими группами).

Модель случайных эффектов расценивает группы как образец реальной или гипотетической популяции групп. Индивидуальные группы не имеют фиксированной доли: они приняты, чтобы представлять множество подобных групп с различиями между ними, приписываемыми случайной вариации или другому «установленному» фактору типа пола, возраста и т.д. Модель случайных эффектов с двумя уровнями отличается от модели, которая не берёт во внимание вложенные группы, хотя обе модели используют случайную или необъяснённую ошибку как части вариации между уровнями первой шкалы (внутригрупповая дисперсия $\sigma^2_{вг}$). Модель случайных эффектов также включает случайную ошибку из-за вариации между группами ($\sigma^2_{мг}$). Дисперсия индивидуального наблюдения в этой модели случайных эффектов равна сумме этих двух компонентов, т.е. $\sigma^2_{вг}+\sigma^2_{мг}$.

Специфические модели

Когда результирующая переменная (зависимая переменная) *y* — числовая и существует единственная значимая объясняющая переменная (предиктор) *x*, то обычная линейная модель со случайными эффектами на двух уровнях проявляется в значении случайного пересечения (свободного члена), которое предполага-

[1] Goldstein H. Multilevel Statistical Models — 3rd edn — Kendall Library of Statistics 3, Arnold, 2003.

ет, что есть линейные отношения между y и x в каждой группе, со своими внутригрупповыми линиями регрессии, имеющими общий наклон β, но различные пересечения (свободные члены) (рис. 42-1, а). Усреднённая линия регрессии имеет такой же наклон и свободный член, равный α, среднее пересечение, усреднённое по всем группам. Случайная ошибка (остаток) для каждой группы — величина, равная расстоянию по вертикали до линии регрессии данной группы, отличная от расстояния до линии с усреднённым свободным членом (см. рис. 42-1, а). Совокупность остатков должна иметь нормальное распределение с нулевым средним и дисперсией, равной σ^2. В пределах каждой группы остатки для групп уровня первой шкалы также должны иметь нормальное распределение с нулевым средним и той же самой дисперсией, равной σ^2. Если размеры групп близки, можно использовать стандартный подход к проверке нормальности и равенства дисперсий для групп шкалы первого уровня, проверяя нормальность с помощью гистограмм остатков, а также используя двухмерные графики, одна из осей которых есть ось остатков, а вторая ось — значение предсказанных величин зависимой переменной.

Эта модель может быть модифицирована различными путями (см. также табл. 42-1), например, разрешив изменить наклон β, изменив случайный набор групп. Модель тогда называется моделью случайного наклона, когда внутригрупповые линии регрессии не параллельны усреднённой линии регрессии (рис. 42-1, б).

Оценка эффекта группирования

Эффект группирования может быть оценён следующими способами.

Вычислением внутригруппового коэффициента корреляции (ВКК, иногда обозначаемый ρ — см. также главу 39), который в структуре с двумя уровнями представляет собой корреляцию между двумя случайно выбранными уровнями шкалы первого уровня в одной случайно выбранной группе.

$$\text{ВКК} = \frac{\sigma^2_{\text{ВГ}}}{\sigma^2 + \sigma^2_{\text{ВГ}}}.$$

ВКК выражает вариацию между группами как пропорцию полной вариации; часто он выражается в виде процентов. ВКК=1, когда нет никакой изменчивости в пределах группы и вся вариация объясняется различиями между группами. ВКК=0, когда нет никакой вариации между группами. Мы можем использовать ВКК, чтобы оценить субъективность решения о важности эффекта группирования.

Сравнением двух моделей, где одна — полная модель случайных эффектов и другая — модель регрессии с тем же самым предиктором (предикторами), но принимающая во внимание наличие группировки. Для проверки различия двух моделей используют соответствующий статистический тест отношения правдоподобия, который подчиняется распределению Хи-квадрата с одной степенью свободы.

ОБОБЩЁННЫЕ ОЦЕНКИ УРАВНЕНИЙ

В концепции[1] обобщённых оценок уравнения (ООУ) мы корректируем оба оцениваемых параметра обобщённых линейных моделей и их стандартные ошибки, принимая во внимание группировку данных в структуре с двумя уровнями. Мы делаем предположения о распределении зависимой переменной, но в отличие от модели случайных эффектов не предполагаем, что межгрупповые остатки имеют нормальное распределение. Расцениваем наличие группировки скорее как

(а) y

Линия регрессии для i-й группы
с пересечением = α_i
и угловым коэффициентом = β

Усреднённая линия
регрессии с пересечением = α
и угловым
коэффициентом = β

α_i

α

0 x

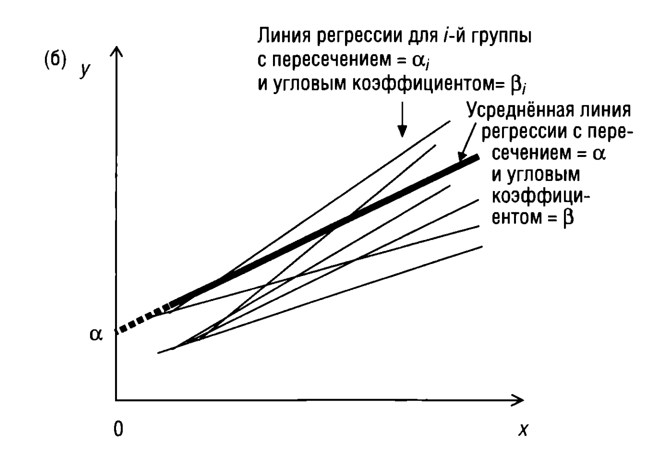


Рис. 42-1. Двухуровневый случайный в модели линейной регрессии с единственным предиктором x.
а) Модель случайного пересечения. Жирная линия представляет усреднённую линию регрессии для всех групп, а каждая из более тонких линий представляет линию регрессии для отдельной группы. Пересечение из i-той группы для определенной линия регрессии отличается от такового усреднённой линии остатком = $\alpha_i - \alpha$, где эти остатки должны иметь нормальное распределение с нулевым средним и дисперсией $\sigma_{\text{Вг}}^2$. Каждая линия регрессии имеет одинаковый наклон с коэффициентом β.
б) Модель случайных наклонов. Жирная линия представляет усреднённую линию регрессии по всем группам, а каждая из более тонких линий представляет линию регрессии для отдельной группы. Пересечение для i-той группы определено линией регрессии, отличающейся от пересечения усреднённой линии регрессии на величину $\alpha_i - \alpha$, и наклон для i-той линии регрессии отличается от наклона усреднённой линии регрессии на остаток $\beta_i - \beta$, где остатки должны иметь нормальное распределение с нулевым средним и дисперсиями $\sigma_{\text{Вг}}^2 + \sigma_{\text{Мг}}^2$ соответственно.

[1] Liang K.Y., Zeger S.L. Longitudinal data analysis using generalized linear models Biometrika, 1986. — Vol. 73. — P. 13–22.

Таблица 42-1. Главные преимущества и неудобства регрессионных методов для анализа сгруппированных данных

Метод	Преимущества	Неудобства
Анализ с агрегированием уровней	Простота	Не учитывает эффекты ковариат для шкалы первого уровня
	Легко выполнять с базовым программным обеспечением	Игнорирует различия в размерах групп и в точности оценок каждого параметра, описывающего группу
		Не всегда может находить соответствующий итоговый показатель
Устойчивые стандартные ошибки, которые учитывают наличие группировок	Относительная простота	Ненадёжно, если число групп велико, например, более 30
	Может включать ковариаты, которые изменяют уровни первой шкалы	Не корректирует оценки параметров для группировок
	Учитывает различия объёмов выборок групп для шкалы первого уровня	
Модель случайных эффектов	Явно учитывает наличие группировок включением межгрупповых и внутригрупповых вариаций в модель	Ненадежно, если нет достаточного количества групп
	Групповые оценки представляют общую информацию от всех групп	Оценки параметров часто смещённые
	Корректирует оценки параметров, стандартные ошибки, доверительный интервал и *P*-значения, чтобы принять во внимание наличие группировки	Для углублённого моделирования требуется наличие навыков
	Может включать ковариаты, которые изменяют уровни первой шкалы	Оценка параметров в модели случайных эффектов для логистической регрессии проблематична
	Учитывает различия объёмов выборок групп для шкалы первого уровня	
	Может расширять иерархию группировки от двух уровней до большего количества	
	Может подгонять различные формы ОЛМ (включая пуассоновскую регрессию)	
ООУ	Относительная простота	Ненадёжно, если число групп велико, например, более 30
	Отсутствие предположений о распределении случайных эффектов (необходимых для кластеров)	Трактовка группировок — досадный раздражитель, вызывающий несвойственный интерес
	Может включать ковариаты, которые изменяют уровни первой шкалы	Требует спецификации рабочей структуры корреляций
	Учитывает различия объёмов выборок групп для шкалы первого уровня	Оценки параметров группы усредняются и не связаны с отдельными наблюдениями в популяции*
	Корректирует оценки параметров, стандартные ошибки, доверительный интервал и *P*-значения, чтобы принять во внимание наличие группировки	

* Эти моменты иногда расценивают и как преимущества в зависимости от заинтересованности.

неудобство, чем интересующее нас свойство, и продолжаем анализ, постулируя «рабочую» структуру для корреляций между наблюдениями в пределах каждой группы. Это не вполне корректно, если достаточно групп с устойчивыми стандартными ошибками и оценки параметров удовлетворительны. Однако мы получим лучшие оценки параметров, если структура правдоподобна. Обычно принимают взаимную структуру корреляции, которая предполагает, что обмен двух уровней в шкале первого уровня в пределах группы не будет изменять оценку.

Подход ООУ иногда называют популяционным усреднением (что касается популяции групп) или маргинальным, потому что оценки параметров представляют эффекты, усреднённые поперёк групп (даже при том, что вся шкала первого уровня включена в анализ). Подход ООУ часто предпочитают более сложному анализу модели случайных эффектов для логистической регрессии (см. главу 30) и иногда пуассоновской регрессии, даже при том, что взаимная структура корреляции, как известно, не корректна в этих ситуациях.

ПРИМЕР

Данные, касающиеся околозубной болезни, были получены на 96 белых мужчинах, инженерах-стажёрах, в возрасте между 16 и 20 годами для обучения в школе Королевских Воздушных Сил, Гальтон, Англия. Каждый из 28 зубов (исключая зубы мудрости) во рту каждого стажёра был исследован на четырёх участках (медиально-щёчный, медиально-язычный, дистально-щёчный и дистально-язычковый). Чтобы упростить анализ, мы рассмотрели подвыборки данных, а именно только (1) медиально-щёчный участок в каждом зубе; это ведёт к структуре с двумя уровнями зубов в пределах субъектов (каждый субъект образует группу), и (2) две интересующие нас переменные; потеря линии сцепления (потеря, измеренная в мм) между зубом и челюстной костью, оценённая на медиально-щёчном участке, и курение стажёра в настоящий момент (да=1, нет=0). Мы хотим оценить, является ли курение фактором риска для болезни дёсен (где большая потеря линии сцепления указывает ухудшение болезни).

В табл. 42-2 показаны итоговые результаты анализа использования различных регрессий, в которых результирующей переменной является потеря линии сцепления (мм) и ковариата — курение. Полная компьютерная распечатка даётся в приложении В. Оценки коэффициентов регрессии для курения и/или их стандартных ошибок изменяются согласно типу выполненного анализа. Два МНК исследования имеют идентичные оценки коэффициентов регрессии (у которых большие величины, чем в трёх других исследованиях), но их стандартные ошибки различны. Стандартная ошибка оценки коэффициента регрессии в МНК-анализе, который игнорирует группировку, существенно меньшая, чем стандартные ошибки в других четырёх исследованиях, т.е. анализ, игнорирующий группировку, даёт результаты с недооценкой стандартной ошибки коэффициента регрессии, а следовательно, слишком узкий доверительный интервал и уменьшенное *P*-значение (уровень значимости). Внутригрупповой коэффициент корреляции в модели случайных эффектов оценён как 0,224. Таким образом, приблизительно 22% вариации в потере линии сцепления после принятия во внимание курения были скорее между стажёрами, чем в пределах стажёров.

В этом специфическом примере мы заключаем из всех пяти исследований, что курение значимо не связано с потерей линии сцепления. Это отсутствие значимости для курения — неожиданное обнаружение и может объясняться тем, что эти стажёры были очень молоды, так что среди них не было курильщиков с длительным стажем.

Таблица 42-2. Итоговые результаты анализа использования различных регрессий, в которых результирующая переменная — потеря линии сцепления (мм)

Вид анализа	Оценки коэффициентов (по курению)	Стандартные ошибки (SE)	95% доверительные интервалы для коэффициентов	Тестовая статистика*	*P*-значение
МНК-регрессия без учёта группировки	−0,0105	0,0235	−0,057 до 0,036	t=−0,45	0,655
МНК-регрессия с устойчивой SE	−0,0105	0,0526	−0,115 до 0,094	t=−0,2	0,842
Агрегированный анализ (МНК-регрессия по групповым средним)	−0,0046	0,0612	−0,126 до 0,117	t=−0,07	0,941
Модель случайных эффектов	−0,0053	0,0607	−0,124 до 0,114	z=−0,09	0,93
ООУ с устойчивыми SE и SES и взаимозаменяемой структурой корреляций	−0,0053	0,0527	−0,108 до 0,098	z=−0,1	0,92

* t=тестовая статистика, отвечающая t-распределению; z= статистика Вальда, отвечающая стандартному нормальному распределению.

МНК= стандартный метод наименьших квадратов.

Данные любезно предоставлены Dr Gareth Griffiths, Dept of Periodontology, Eastman Dental Institute, University College London. UK.

43 Систематические обзоры и метаанализ

СИСТЕМАТИЧЕСКИЙ ОБЗОР

Что это такое?

Систематический обзор[1] — оформленный надлежащим образом и точный процесс соединения информации всех относящихся к делу исследований (как опубликованных, так и неопубликованных) по определённому состоянию здоровья; обычно это клинические исследования того же или подобного лечения, но могут быть и наблюдательные. Ясно, что систематический обзор — интегральная часть ДМ, которая применяет результаты лучших имеющихся данных вместе с клинической экспертизой для лечения и выхаживания пациентов. Его роль так важна в ДМ, что он стал центром международной сети клиницистов, методологов и потребителей, которые образовали Кокрановское сотрудничество. Это привело к созданию Кокрановской библиотеки, содержащей регулярно обновляемые базы данных по здравоохранению на основе научно-доказательных материалов, включая Кокрановскую базу данных Систематических Обзоров. Полный доступ к этим обзорам требует подписки, но резюме свободно доступны в Интернете (www.cochrane.org/reviews).

Что этим достигается?

- Уточнение и сокращение. Большие количества информации уточняются и сокращаются до приемлемого размера.
- Эффективность. Выполнить систематический обзор быстрее и дешевле, чем провести новое исследование. Он может предостеречь других от ненужных исследований и может сократить разрыв во времени от медицинских исследований до их внедрения.
- Способность к обобщению и согласованность. Результаты часто можно распространить на более широкую популяцию пациентов на более широком фоне, чем это было бы возможно при единичном исследовании. Можно оценить согласованность в результатах различных исследований и определить несогласованности.
- Надёжность. Систематический обзор уменьшает число ошибок и таким образом повышает надёжность и точность рекомендаций по сравнению со случайными обзорами или единичными исследованиями.
- Мощность и точность. Количественный систематический обзор (см. «Метаанализ») имеет большую мощность для обнаружения интересующих эффектов и даёт им более точные оценки, чем одиночное исследование.

МЕТААНАЛИЗ

Что это такое?

Метаанализ, или обзор, — особая форма систематического обзора, который фокусируется на числовых результатах. Основная цель метаанализа — соединить результаты от отдельных исследований, чтобы дать, если это уместно, оценку общего или среднего интересующего нас эффекта (например, ОР). Направление и величину этого среднего эффекта вместе с рассмотрением соответствующего доверительного интервала и результата проверки гипотез можно использовать для принятия решений об исследуемом лечении и лечении пациентов.

Статистический подход

1. **Мы делаем выбор относительно интересующего нас эффекта и, если есть исходные данные, оцениваем его для каждого исследования**. Однако на практике приходится извлекать эти эффекты из опубликованных результатов. Если исход клинического исследования, сравнивающего два метода лечения:

- числовой — эффект может быть обусловлен различием средств лечения. Нулевая разность означает отсутствие эффекта лечения;
- бинарный (например, умер/живой) — мы рассматриваем риски исхода (например, смерть) в группах лечения. Эффектом может быть разница в рисках или в их отношении, ОР. Если разность рисков равна нулю или ОР=1, эффекта лечения нет.

2. **Проверяем статистическую однородность и получаем оценку статистической неоднородности**. Мы имеем статистическую разнородность, когда есть реальная вариация между интересующими нас эффектами и остальными причинами. Можно выполнять проверку гипотезы гомогенности, чтобы исследовать, совместимы ли вариации в индивидуальных эффектах со случайным эффектом. Однако это испытание имеет низкую мощность, чтобы обнаружить разнородность, если есть немного исследований в метаанализе, и может, наоборот, давать высоко значимый результат, если оно включает много больших исследований, и даже когда имеется разнородность, вряд ли она скажется на выводах. Индекс I^2, который не зависит от количества исследований, типа данных, по которым получен результат, или выбора эффекта лечения (например, относительный риск), можно использовать, чтобы определить величину влияния разнородности и оценить противоречивость (см. пример). I^2 представляет собой долю от полной вариации по всем исследованиям и изменяет свою величину от 0 до 100%. Значение 0% указывает на отсутствие наблюдаемой разнородности. Если есть доказательства статистической разнородности, мы должны действовать осторожно, исследуя возможные причины её присутствия, и изменить соответственно наш подход, возможно, разделяя исследования на подгруппы с подобными характеристиками.

3. **Оценить среднюю величину интересующего нас эффекта (с доверительным интервалом) и выполнить проверки гипотез по этому эффекту подходящим критерием (например, что истинный ОР=1)**. В этом контексте вы можете прийти к моделям с «фиксированными эффек-

[1] Chalmers I., Altman D.G. (eds.) Systematic Reviews // British Medical Journal Publishing Group, London, 1995.

тами» и «случайными эффектами». Если нет никакого свидетельства статистической разнородности, используйте модель фиксированных эффектов (которая принимает значение типа лечения одним и тем же в каждом исследовании, и любая наблюдаемая вариация в этом случае объясняется случайностью выборки); в противном случае используем модель случайных эффектов (она предполагает, что отдельные исследования представляют собой случайную выборку из популяции, в которой имеется средний эффект лечения и в которой изменяются индивидуальные эффекты исследования).

4. **Интерпретировать результаты и представить выводы**. Полезно суммировать результаты от каждого исследования (например, размер выборки, базовые характеристики, интересующий эффект, такой, как ОР и соответствующие ДИ) в таблице (см. пример). Наиболее частое графическое изображение — так называемый форест-графика (рис. 43-1), в котором оценённый эффект (с ДИ для каждого исследования) и их среднее отмечены по длине вертикальной линией, представляющей «эффект без лечения» (т.е. эта линия соответствует значению «единица», если эффект — ОР). Вначале мы проверяем, располагаются ли оценённые эффекты от различных исследований на одной и той же стороне от этой линии. Затем можно использовать ДИ, чтобы удостовериться, совместимы ли результаты (если ДИ перекрываются), с целью определения, можно ли объяснить несовместимые результаты небольшими размерами выборки (если ДИ широкие) и оценки значимости индивидуальных и общих эффектов (путём наблюдения, пересекает ли вертикальная линия некоторые или все ДИ).

Преимущества и недостатки

Поскольку метаанализ — особая форма систематического обзора, он имеет все преимущества последнего (см. «Что этим достигается?»). В частности, метаанализ из-за своего увеличенного размера выборки способен обнаружить эффекты лечения с большей мощностью и оценить эти эффекты с большей точностью, чем любое единичное исследование. Его преимущества, вместе с введением компьютерной программы метаанализа, привели эту форму обзора к распространению. Однако некорректное применение может вести к ошибочным заключениям относительно эффективности лечения. Следующие основные проблемы должны быть тщательно исследованы и разрешены до того, как проводить метаанализ.

- Публикационное смещение — тенденция включать в анализ только результаты из опубликованных статей; это способствует статистически значимым выводам.
- Клиническая гетерогенность — разница в популяции пациентов, исходных измерениях, определении переменных и/или длительности наблюдения исследований, включённых в анализ, создают проблемы несовместимости.
- Различие в качестве — качество разработки и проведения исследований может варьировать. Хотя решение этой дилеммы — придание большего веса лучшим исследованиям, любую систему оценки можно критиковать на основании того, что она произвольна.
- Зависимость — результаты исследований, включённых в анализ, могут не быть независимыми, т.е. когда результаты исследования публикуются более чем по одному случаю.

ПРИМЕР

Пациента с тяжёлой стенокардией обычно направляют на перкутанную транслюминальную коронарную ангиопластику или операцию шунтирования коронарной артерии (CABG). Результаты восьми опубликованных рандомизированных исследований были собраны в метаанализ из 3371 пациента (1661 CABG, 1710 PTCA). Основные свойства этих исследований показаны в табл. 43-1. Результаты для составной конечной точки, сердечная смерть плюс несмертельный ИМ в первый год наблюдения показаны на рис. 43-1. Оценённые ОР даны для группы PTCA по сравнению с группой CABG. На рисунке использована логарифмическая шкала для ОР, чтобы достичь симметричных ДИ. Хотя индивидуальные оценки ОР варьируют довольно значительно, от уменьшения риска до его довольно большого увеличения, все доверительные интервалы до некоторой степени перекрываются.

Более формальную оценку неоднородности обеспечивает критерий однородности Хи-квадрат Кокрана, она даёт незначимый результат (значение Q-статистики Кокрана равно 13,2 при степени свободы df=8−1=7, P=0,07). Однако $I^2=100\times(Q-df)/Q=100\times(13{,}2-7)/13{,}2=47\%$ (95% ДИ от 0 до 76%) предполагает умеренную несогласованность в целом по всем исследованиям и указывает на необходимость осторожной интерпретации объединённой оценки относительного риска для всех исследований. ОР=1,03 для всех комбинированных исследований (95% ДИ от 0,79 до 1,50) означает, что нет оснований для различия между этими двумя стратегиями реваскуляризации. Интересно отметить, что в ранний период наблюдения распространённость стенокардии была выше у пациентов с PTCA, чем с CABG.

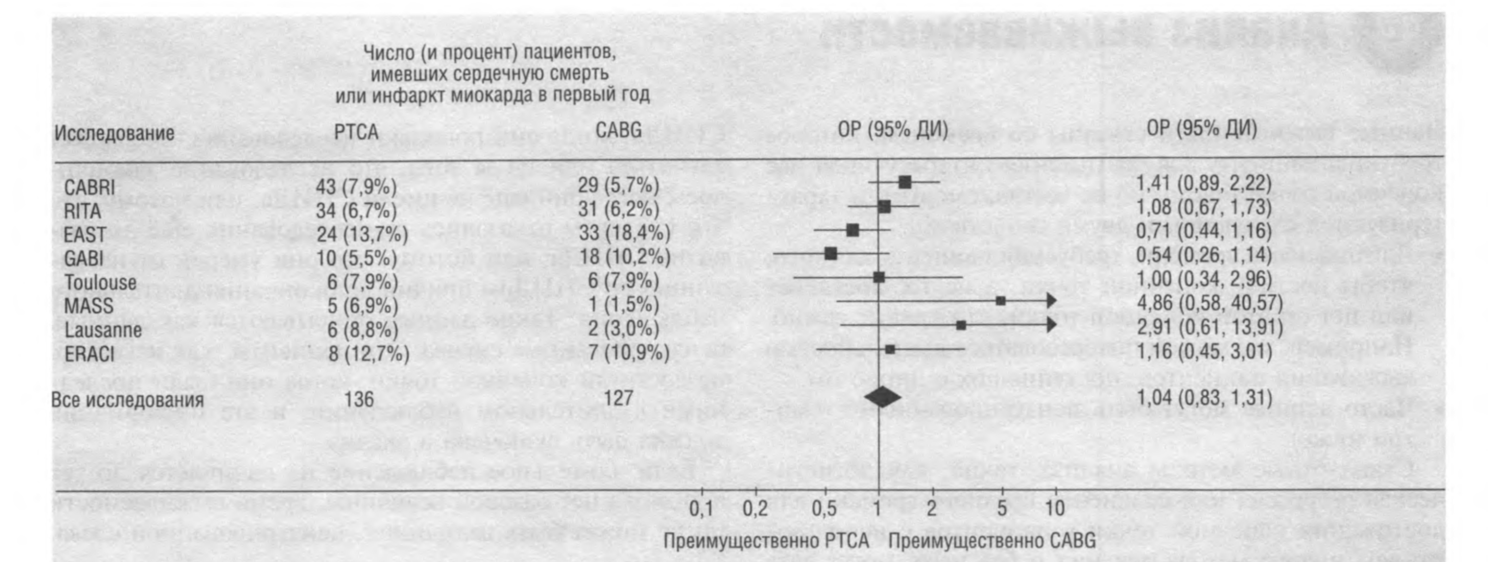


Рис. 43-1. Форест-график относительных рисков (95% доверительный интервал) сердечных смертей или инфаркта миокарда для группы перкутанной транслюминальной коронарной ангиопластики по сравнению с группой операции шунтирования коронарной артерии в первый год от рандомизации. PNCA — перкутанная транслюминальная коронарная ангиопластика; CABG — операция шунтирования коронарной артерии

Таблица 43-1. Характеристики восьми рандомизированных исследований, сравнивающих перкутанную транслюминальную коронарную ангиопластику и шунтирование коронарной артерии

	Страна	Главный исследователь	Одно- или многососудистое	Число пациентов		Период наблюдения, лет
				CABG	PTCA	
Coronary Angioplasty Bypass Revascularisation Investigation(CABRI)	Европа	A.F. Rickards	Много	513	541	1
Randomised Intervention on Treatment of Angina Trial (RITA)	Англия	J.R. Hampton	Одно (n=456) и много (n=555)	501	510	4,7
Emory Angioplasty versus Surgery Trial (EAST)	США	S.B. King	Много	194	198	3+
German Angioplasty Bypass Surgery Investigation (GABI)	Германия	C.W. Hamm	Много	177	182	1
The Toulouse Trial (Toulouse)	Франция	J. Puel	Много	76	76	2,8
Medicine Angioplasty orSurgery study (MASS)	Бразилия	W. Hueb	Одно	70	72	3,2
The Lausanne trial (Lausanne)	Швейцария	J.-J. Goy	Одно	66	68	3,2
Argentine Trial of PTCA versusCABG (ERACI)	Аргентина	A. Rodriguez	Много	64	63	3,8

Адаптировано из Pocock S.J., Henderson R.A., Rickards A.F., el al. A metaanalysis of randomised trials comparing coronary angioplastywith bypass surgery. Lancet, 1995. — Vol. 346. — P. 1184–1189, with permission. PNCA — перкутанная транслюминальная коронарная ангиопластика; CABG — операция шунтирования коронарной артерии.

44 Анализ выживаемости

Данные выживаемости связаны со временем, которое требуется пациенту для достижения интересующей нас конечной точки (часто, но не всегда, смерти), и характеризуются следующими двумя свойствами.

- Длительность времени, требуемая пациенту для того, чтобы достичь конечной точки, а не то, достигнет или нет он этой конечной точки, что крайне важно. Например, мы можем интересоваться длительностью выживания пациентов, поступивших с циррозом.
- Часто данные могут быть цензурированными (смотри ниже).

Стандартные методы анализа, такие, как логистическая регрессия или сравнение среднего времени для достижения конечной точки у пациентов с использованием нового метода лечения и без него, могут дать сбивающие с толку результаты из-за цензурированных данных. Для таких ситуаций разработан ряд статистических способов, известных как **методы выживаемости**[1].

ЦЕНЗУРИРОВАННЫЕ ДАННЫЕ

Время выживаемости рассчитывается от некоторой базовой величины, которая отражает естественную «точку старта» для исследования (например, время операции или диагноз условия) до того времени, когда пациент достигает интересующую нас конечную точку. Однако часто мы можем не знать, когда пациент достиг конечной точки, зная только то, что он за время исследования не достиг такой конечной точки. Например, пациенты в исследовании нового лекарства для инфекции вирусом иммунодефицита человека могут не иметь СПИДа, когда они покидают исследование. Это может случиться или из-за того, что исследование закончилось, когда они ещё не имели СПИДа, или потому, что эти субъекты отказались от исследования ещё до развития СПИДа, или потому, что они умерли от несвязанных со СПИДом причин до окончания длительного наблюдения. Такие данные описываются как данные, цензурированные справа. Эти пациенты, как известно, не достигли конечной точки, когда они были последними в длительном наблюдении, и эта информация должна быть включена в анализ.

Если длительное наблюдение не начинается до тех пор, пока нет базовой величины, время выживаемости также может быть величиной, цензурированной слева.

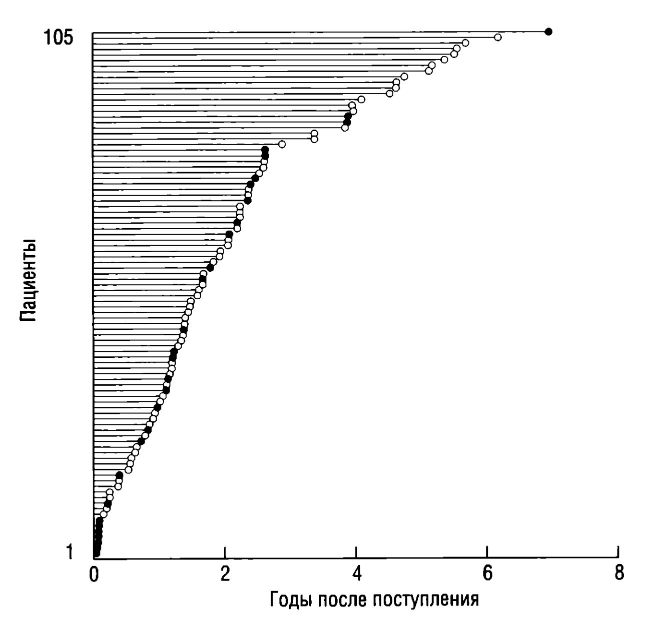


Рис. 44-1. Исследование выживаемости 105 пациентов после поступления с циррозом. Чёрные кружки означают пациентов, которые умерли, белые — тех, кто остался жив в конце срока наблюдения.

ОТОБРАЖЕНИЕ ДАННЫХ ВЫЖИВАЕМОСТИ

- Для каждого пациента может быть начерчена отдельная горизонтальная линия, причём её длина показывает время выживаемости. Линии нанесены слева направо, и пациенты, которые достигают конечной точки и которые цензурированы, могут различаться с помощью символов в конце линии (рис. 44-1). Однако эти графики не суммируют данные и трудно освоиться с опытом выживаемости в целом.
- Кривые выживаемости, обычно рассчитываемые методом Каплана–Майера, отображают кумулятивную (накопленную) вероятность (вероятность выживаемости) пациента, остающегося без конечной точки в какое-то время после базового (рис. 44-2). Вероятность выживаемости будет изменяться только тогда, когда наступает конечная точка, и таким образом получившаяся в результате «кривая» наносится как ряд этапов. Альтернативный метод расчёта вероятностей выживаемости, применяющий подход таблиц продолжительности жизни, можно использовать только тогда, когда известно, что время для

Кумулятивный процент выживавмости

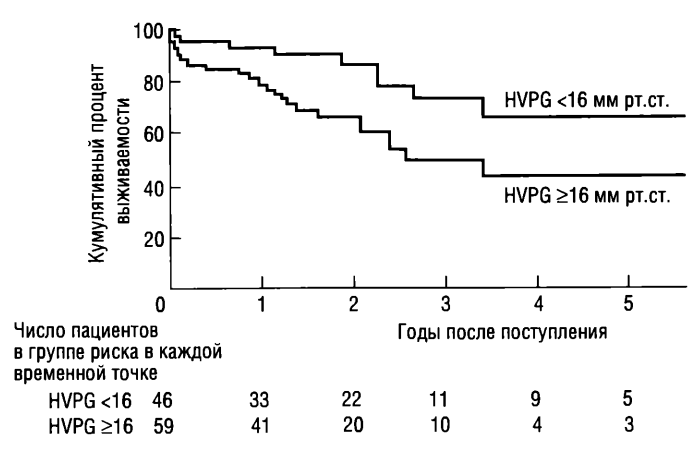


Рис. 44-2. Кривые Каплана–Майера, показывающие вероятность выживаемости, выраженную в процентах, после поступления с циррозом, стратифицированные измерением базового уровня давления в воротной вене.

[1] Colletl D. Modelling Survival Dam in Medical Research. — 2nd edn. — Chapman unil Hall // CRC, London, 2003.

достижения конечной точки находится в пределе определённого временного интервала (например, в пределе одного года). Вероятности выживаемости, применяющие другой метод, просты, но требуют времени для расчёта и могут быть легко получены из большинства пакетов компьютерных программ.

СУММИРУЮЩЕЕ ВЫЖИВАНИЕ

Мы часто суммируем выживание, указывая вероятности выживаемости (с доверительными интервалами) в определённых временных точках на кривой, например, уровень пятилетней выживаемости у пациентов после лечения рака груди. Также можно определить среднее время, требуемое для достижения конечной точки (время, за которое 50% субъектов улучшили своё состояние).

СРАВНЕНИЕ ВЫЖИВАЕМОСТИ

Мы можем захотеть оценить воздействие ряда интересующих нас факторов на выживаемость, например, лечения, тяжести заболевания. Кривые выживаемости можно нанести отдельно для подгрупп пациентов; они дают средние, чтобы визуально оценить, с различной ли скоростью достигают группы пациентов конечной точки (рис. 44-2). Мы можем формально проверить, имеется ли какая-либо значимая разница в частоте прогрессирования между различными группами, например, применяя лог-ранговый критерий или модели регрессии.

Лог-ранговый критерий

Этот непараметрический критерий обращается к нулевой гипотезе о том, что в исследуемых группах нет разницы во времени выживаемости, и сравнивает события, происходящие во всех временных точках на кривой выживаемости. Мы можем оценить влияния независимой роли более чем одного фактора на время до конечной точки, применяя лог-ранговый критерий.

Модели регрессии

Можно генерировать регрессионную модель для количественного расчёта соотношений между одним или более интересующими нас факторами и выживаемостью. В каждой точке во время t субъект i имеет мгновенный риск достижения конечной точки (часто известный как риск или $\lambda_i(t)$, при условии, что он не достиг её вплоть до этой точки во времени). Например, если конечная точка — смерть, риск — это риск умереть во время t. Этот мгновенный риск обычно очень мал и имеет ограниченный интерес. Однако можно узнать, имеется ли какая-либо систематическая разница между этими рисками во всех временных точках для пациентов с различными характеристиками. Например, понижен ли риск в целом у пациентов, пролеченных новым терапевтическим подходом, по сравнению с пациентами, пролеченными плацебо, если принимать во внимание и другие факторы, такие, как возраст или тяжесть заболевания?

Можно применить регрессионную модель Кокса пропорционального риска для проверки независимых эффектов ряда объясняющих переменных (факторов) на риски. Она имеет вид:

$$\lambda_i(t)=\lambda_0(t)\exp\{\beta_1 x_1+ \beta_2 x_2+ \ldots + \beta_k x_k\},$$

где $\lambda_i(t)$ — риск для пациента i во время t, $\lambda_0(\mathrm{t})$ — произвольный базовый риск (в котором мы не заинтересованы), x_1, .., x_k — объясняющие переменные (предикторы) в модели и β_1, .., β_k —соответствующие коэффициенты.

Мы получаем оценки b_1, .., b_k этих параметров, используя разновидность максимального правдоподобия, известную как частное правдоподобие. Экспонента от этих значений (например, $\exp\{b_1\}=e^{b_1}$) есть оценка относительного риска или отношения рисков заболевания. Для конкретного значения «x_1» относительный риск равен повышению или понижению опасности достижения конечной точки, связанной с единичным увеличением до (x_1+1), с учётом других предикторов в модели. Относительный риск интерпретируется подобным же образом для отношения шансов в логистической регрессии или как отношение интенсивностей в пуассоновской регрессии, поэтому величины выше единицы означают повышенный риск, величины ниже единицы — пониженный риск и равные единице означают, что нет повышенного или пониженного риска достижения конечной точки. Доверительный интервал можно рассчитать для относительного риска и использовать критерий значимости для оценки его отклонения от 1.

В этой модели допускается постоянный в течение времени относительный риск (т.е. предполагается, что риски для сравниваемых групп пропорциональны). Важно проверить это допущение, или применяя графические методы, или включая взаимодействие между ковариатой и log(время) в модель и гарантируя, что оно несущественно.

Другие модели также могут быть использованы для описания данных выживаемости, например, экспоненциальная модель или модель Вейбулла. Однако они — вне рамок этой книги.

ПРИМЕР

Уровень давления в воротной вене (УДВВ), как известно, связан с тяжестью алкогольного цирроза, но редко используется как предиктор выживаемости пациентов с этим заболеванием. Чтобы оценить клиническую важность этого измерения, 105 пациентов, поступивших с циррозом и подвергшихся печёночной венографии, исследовались в среднем 566 дней. Наблюдение за выживаемостью этих пациентов показано на рис. 44-1. За период наблюдения 33 пациента умерли. Кривые Каплана–Майера, показывающие кумулятивную частоту выживаемости в каждой точке времени после базовой, изображены отдельно для тех субъектов, у которых УДВВ было менее 16 мм рт.ст. (значение, предположительно дающее прогностическую значимость), и для тех субъектов, у которых УДВВ был 16 мм рт.ст. или выше (см. рис. 44-2).

Метод	Хи-квадрат	Степени свободы	*P*
Лог-ранг	5,2995	1	0,0213

Таким образом, имеется существенное (P=0,02) различие между временем выживаемости в этих двух группах. Через три года после поступления в госпиталь 73,1% пациентов, имеющих низкий УДВВ, остались живы в отличие от 49,6%, имеющих более высокие значения (см. рис. 44-2).

Была использована регрессионная модель Кокса пропорционального риска для того, чтобы проверить, можно ли объяснить это соотношение разницей в любых известных прогностических или демографических факторах в исходном состоянии. Для включения в эту модель было рассмотрено 12 переменных, в том числе демографические, клинические и лабораторные показатели. Графические методы дали предположение, что допущение пропорциональных рисков для этих переменных было уместным. Для выбора конечной оптимальной модели использовалась процедура пошагового отбора, и результаты показаны в табл. 44-1.

Результаты в табл. 44-1 показывают, что повышенное УДВВ остаётся независимо связанным с более короткой выживаемостью после корректировки других факторов, которые, как известно, связаны с более плохим исходом. В частности, пациенты с УДВВ в 16 мм рт.ст. или выше имели риск смерти в 2,46 раза (=exp{0,90}) выше, чем пациенты с более низкими уровнями (P=0,04) после корректировки других факторов. Другими словами, опасность смерти у этих пациентов была увеличена на 146%. Кроме того, повышенное протромбиновое время (риск повышается на 5% на дополнительную секунду), повышенный уровень билирубина (риск повышается на 5% на 10 дополнительных ммоль/л), наличие асцита (риск увеличивается на 126% при увеличении уровня) и длительное эндоскопическое лечение в анамнезе (риск увеличивается на 246%) — все они независимо и значимо связаны с исходом.

Таблица 44-1. Результаты регрессионной модели Кокса пропорционального риска

Переменная (и код группы)	Степени свободы	Оценка параметра	Стандартная ошибка оценки	Достигнутый уровень значимости *P*	Относительный риск	95% ДИ относительного риска
УДВВ* (0= <16,1=≥16 мм рт.ст.)	1	0,90	0,44	0,04	2,46	(1,03 — 5,85)
Протромбиновое время, с	1	0,05	0,01	0,0002	1,05	(1,02 — 1,07)
Билирубин (10 ммоль/л)	1	0,05	0,02	0,04	1,05	1,00 — 1,10)
Асцит (0 — нет, 1 — средний, 2 — умеренный/тяжёлый)	1	0,82	0,18	0,0001	2,26	(1,58 — 3,24)
Длительное эндоскопическое лечение в анамнезе (0 — нет, 1— да)	1	1,24	0,41	0,003	3,46	(1,54 — 7,76)

* Уровень давления в воротной вене.

Данные любезно предоставлены Dr D. Patch and Dr A.K. Burroughs, Liver Unit, Royal Free Hospital, London, UK.

45 Байесовские методы

ЧАСТОТНЫЙ ПОДХОД

Критерии проверки гипотез, описанные в этой книге, основаны на частотном подходе к вероятности и подразумевают, что этот подход принимает во внимание, сколько раз должно произойти явление, если бы мы вынуждены были повторять эксперимент много раз. Этот подход иногда подвергается критике по следующим причинам.

- Он использует только информацию, полученную от настоящего исследования, и не включает в процесс любую другую об интересующем нас эффекте, т.е. взгляды клиницистов об относительной эффективности двух терапий до того, как предпринято клиническое исследование.
- Он не обращается непосредственно к вопросам наибольшего интереса. При сравнении препаратов мы фактически хотим знать, какой из них более эффективен. Однако частотный подход проверяет гипотезу, что эти два препарата эффективны одинаково. Хотя мы заключаем, что один препарат превосходит другой, если P небольшое, эта вероятность (т.е. значение P) описывает шанс получения наблюдаемых результатов, если препараты равно эффективны, а не шанс, что один препарат более эффективен, чем другой (наш реальный интерес).
- Он имеет тенденцию переоценивать роль критерия проверки гипотез и то, значим ли результат, а не смысл результатов.

БАЙЕСОВСКИЙ ПОДХОД

Альтернативный байесовский метод вывода отражает персональную степень доверия исследователя к гипотезе, возможно, на основании уже имеющейся информации. Исследователи обычно различаются по своей степени доверия к гипотезе; кроме того, это доверие может изменяться по мере поступления новой информации. Байесовский подход рассчитывает вероятность того, что гипотеза истинна (наш центр интереса), путём обновления предшествующих мнений о гипотезе, по мере того как новые данные становятся доступными.

Условная вероятность

В байесовских анализах фундаментален особый тип вероятности, известный как условная вероятность. Это вероятность события, при условии, что другое событие уже произошло. В качестве иллюстрации рассмотрим пример. Частота гемофилии А в общей популяции — около 1 на 10 тыс рождений мальчиков. Однако, если мы знаем, что носитель гемофилии — женщина, эта частота возрастает почти до 1 на 2 рождения мальчиков. Вероятность того, что мальчик родится с гемофилией, при условии, что его мать — носитель заболевания, очень отличается от безусловной вероятности того, что он родится с гемофилией, если статус его матери неизвестен.

Теорема Байеса

Предположим, что мы исследуем гипотезу (например, что эффект лечения равен некоторой величине). Теорема Байеса преобразует априорную вероятность, описывая доверие исследователя к гипотезе до того, как исследование проведено, в апостериорную вероятность, описывая его доверие после. Апостериорная вероятность — фактически условная вероятность гипотезы, использующей результаты исследования. Теорема Байеса утверждает, что апостериорная вероятность пропорциональна априорной, умноженной на величину, называемую правдоподобием наблюдаемых результатов, которая описывает правдоподобие наблюдаемых результатов, если гипотеза верна.

ДИАГНОСТИЧЕСКИЕ КРИТЕРИИ В БАЙЕСОВСКИХ РАМКАХ

Почти все клиницисты интуитивно применяют байесовский подход при постановке диагноза. Они выстраивают картину болезни пациента на основании клинического анамнеза и/или наличия симптомов и признаков. Исходя из этого они ставят наиболее вероятный диагноз, устраняя другие на том основании, что они маловероятны при имеющихся знаниях о пациенте. Впоследствии они могут подтвердить или изменить этот диагноз в свете новых данных, например, если пациент отвечает на лечение или развились новые симптомы.

Когда пациент посещает клинику, у врача обычно есть некоторые идеи относительно вероятности того, что субъект болен — априорная, или дотестовая вероятность. Если о пациенте больше ничего не известно, это просто распространённость заболевания в популяции. Мы можем применить теорему Байеса для преобразования априорной вероятности в апостериорную. Легче всего этого достичь, если ввести в теорему Байеса отношение правдоподобия, основываясь на информации, полученной из последнего исследования (например, результат диагностического теста). Отношение правдоподобия положительного результата теста — это шанс положительного результата теста, если пациент болен, делённый на то, если он не болен. Мы ввели отношение правдоподобия в главе 38 и показали, что его можно использовать для того, чтобы показать полезность диагностического критерия. В том же контексте теперь мы применяем его для выражения теоремы Байеса в терминах шансов:

Апостериорные шансы заболевания = (априорные шансы)×(отношение правдоподобия положительного результата теста),

где

априорные шансы = априорная вероятность/(1−априорная вероятность)

и

отношение правдоподобия = = чувствительность/(−специфичность).

Апостериорные шансы рассчитать легко, но для более лёгкой интерпретации мы преобразуем шансы обратно в вероятность, применяя взаимоотношение:

Апостериорная вероятность = апостериорные шансы/(1+апостериорные шансы).

Эта апостериорная, или послетестовая вероятность — вероятность того, что пациент имеет заболевание, при условии положительного результата теста. Это подобно прогностичности положительного результата теста, но принимает во внимание априорную вероятность того, что субъект имеет заболевание.

Более простой путь выполнения этих расчётов — применить номограмму Фагана (рис. 45-1); путём соединения точки априорной вероятности (выраженной в процентах) с отношением правдоподобия, продолжая линию, мы можем оценить апостериорную вероятность.

НЕДОСТАТКИ БАЙЕСОВСКИХ МЕТОДОВ

Частью любого байесовского анализа становится необходимость уточнения априорной вероятности гипотезы (т.е. априорную вероятность того, что пациент болен). Из-за субъективной природы предшествующих случаев отдельные исследователи и клиницисты могут выбрать для них различные значения. Байесовские методы часто критикуют как произвольные. Однако там, где наиболее свежие данные из исследования (например, вероятность) очень убедительны, влияние априорной информации сведено до минимума (в своей крайности априорная информация не будет вообще влиять на результаты).

Расчёты, введённые во многие байесовские анализы, сложны, обычно требуют изощрённых пакетов статистических программ. Несмотря на интуитивную притягательность, байесовские методы не находят широкого применения. Однако наличие мощных персональных компьютеров приведёт к более частому их применению.

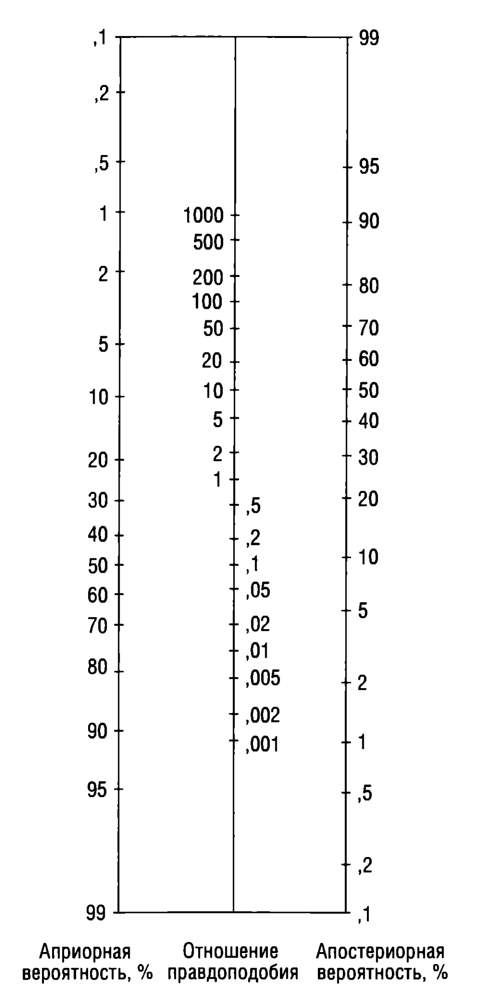


Рис. 45-1. Номограмма Фагана для интерпретации результата диагностического критерия.

ПРИМЕР

На примере в главе 38 мы показали, что у реципиентов костного трансплантата вирусная нагрузка выше $5\times\log_{10}$ genomes/мл дала оптимальную чувствительность и специфичность теста для прогноза развития тяжёлого клинического заболевания. Отношение правдоподобия для положительного результата этого теста по этому значению точки разделения было 13,3.

Если мы верим, что распространённость тяжёлого заболевания в результате цитомегаловирусной (ЦМВ) инфекции после костной трансплантации приблизительно 33%, априорная вероятность тяжёлого заболевания у этих пациентов составляет 0,33.

Априорные шансы =0,33/0,67=0,493.

Апостериорные шансы =0,493×отношение правдоподобия =0,493×13,3=6,557.

Апостериорная вероятность =6,557/(1+6,557)=6,557/7,557=0,868.

Если у пациента ЦМВ нагрузка выше $5\times\log_{10}$ genomes/мл и мы допускаем, что априорная вероятность тяжёлого заболевания равна 0,33, то предполагаем, что пациент имеет 87% шанс развития тяжёлого заболевания. Это можно также прямо оценить из номограммы Фагана (см. рис. 45-1) путём соединения априорной вероятности в 33% с отношением правдоподобия в 13,3 и продлевая линию до пересечения с осью апостериорной вероятности. Наоборот, если мы полагаем, что вероятность того, что пациент получит тяжёлое заболевание только 20% (т.е. априорная вероятность равна 0,2), тогда апостериорная вероятность будет 77%.

В обоих случаях апостериорная вероятность гораздо выше, чем априорная, означая полезность положительного результата теста. Более того, оба результата показывают, что у пациента высокий риск развития тяжёлого заболевания после трансплантации, и это причина, чтобы начинать антиЦМВ терапию. Несмотря на очень различающиеся априорные вероятности общее заключение в каждом случае остаётся одинаковым.

Приложение А. Статистические таблицы

Это приложение содержит статистические таблицы, которые обсуждались в тексте. Мы дали только ограниченные значения *P*, поскольку данные обычно анализируют с помощью компьютера, и эти значения включают в результат анализа. Другие тексты, такие, как тексты Fisher и Yates, содержат более обширные таблицы. Вы можете также получить значения *P* из некоторых пакетов компьютерных программ, задав величину статистики критерия. Пустые клеточки в таблице означают, что такое значение не существует.

Таблица А1 содержит вероятность с двусторонней критической областью распределения переменной *z*, которая подчиняется стандартному нормальному распределению. Величины *P* в таблице А1 относятся к абсолютным значениям *z*, так что, если *z* отрицательно, мы игнорируем его знак. Например, если статистика критерия, которая подчиняется стандартному нормальному распределению, имеет значение 1,1, то *P*=0,271.

Таблица А2 и **таблица А3** содержат вероятность с двусторонней критической областью распределения переменной, которая подчиняется *t*-распределению (таблица А2) или распределению Хи-квадрат Пирсона (таблица А3) с данными степенями свободы (*df*). Чтобы применить таблицу А2 или А3, если абсолютная величина статистики критерия (с данным значением *df*) лежит между табулированными величинами в двух колонках, тогда двусторонняя величина *P* лежит между величинами *P*, определёнными в верхней части этих колонок. Если статистика критерия находится справа от конечной колонки, то *P* <0,001; если она слева от второй колонки, то *P* >0,1. Например, (i) таблица А2: если статистика критерия равна 2,62 с *df*=17, то 0,01 <*P* < 0,05; (ii) таблица А3: если статистика критерия =2,62 с *df*=17, то *P* <0,001.

Таблица А4 содержит часто применяемые величины *P* и соответствующие им величины для *z*, переменной со стандартным нормальным распределением. Эта таблица может использоваться для получения множителей с целью расчёта доверительных интервалов для нормально распределённых переменных. Например, для 95% доверительного интервала множитель равен 1,96.

Таблица А5 содержит величины *P* для переменной, которая подчиняется *F*-распределению с рассчитанными степенями свободы в числителе и знаменателе. При сравнении дисперсий (см. главу 35) мы обычно применяем двустороннюю величину *P*. Для дисперсионного анализа (см. главу 22) мы применяем одностороннюю величину *P*. Для заданных степеней свободы в числителе и знаменателе критерий значим на уровне *P*, упоминаемом в таблице, если статистика критерия больше, чем табулированная величина. Например, если статистика критерия 2,99 с *df*=5 в числителе и *df*=15 в знаменателе, то *P* <0,05 для одностороннего критерия.

Таблица А6 содержит двусторонние *P*-значения знакового критерия для *r* специфических ответов из общего количества ответов *n'*. Для одновыборочного критерия *r* равно числу значений переменной выше (или ниже) медианы. Для парного критерия *r* равно числу положительных (или отрицательных) разностей или числу предпочтений особого лечения (см. главу 23). *n'* либо по усмотрению равно числу значений, не равных медиане, ненулевой разнице или фактическим предпочтениям. Например, если мы наблюдаем три положительных разности из восьми ненулевых разностей, тогда *P*=0,726.

Таблица А7 содержит диапазоны значений, которые определяют верхний и нижний пределы приблизительных 90, 95 и 99% доверительных интервалов для медиан. Например, если размер выборки 23, то границы 95% доверительного интервала определяются 7-м и 17-м упорядоченными значениями.

Для выборки размером более 50 найдите наблюдения, которые соответствуют рангам (до ближнего целого числа), равным: (i) $n/2-z\sqrt{n}/2$ и (ii) $1+n/2+z\sqrt{n}/2$, где *n* — объём выборки и *z*=1,64 для 90% ДИ, *z*=1,96 для 95% ДИ и *z*=2,58 для 99% ДИ (значения *z* получены из стандартного нормального распределения *N*(0;1), таблица А4). Эти наблюдения определяют (i) нижнюю и (ii) верхнюю доверительные границы для медианы.

Таблица А8 содержит диапазон значений для суммы рангов (T_+ или T_-), которая определяет значимость в критерии знаковых рангов Вилкоксона (см. главу 20). Если сумма рангов положительных (T_+) или отрицательных (T_-) разностей находится вне *n'* ненулевых разностей, равна или находится вне табулированных границ, критерий значим при упомянутой величине *P*. К примеру, если имеются 16 ненулевых разностей и T_+=21, тогда 0,01 < *P* < 0,05.

Таблица А9 содержит диапазон значений для суммы рангов (Т), которая определяет значимость для двухвыборочного критерия Вилкоксона ранговых сумм (глава 21) при (а) 5% уровне и (*b*) 1% уровне. Предположим, у нас две выборки размерами n_S и n_L, где $n_S \le n_L$. Если сумма рангов группы с меньшим объёмом выборки, n_S равна или находится вне табулированных границ, критерий значим при (а) 5% уровне или (*b*) 1% уровне. Например, если n_S=6 и n_L=8, а сумма рангов в группе из 6 наблюдений равна 39, то *P* >0,05.

Таблица А10 и **таблица А11** содержат двусторонние *P*-значения для корреляционных коэффициентов Пирсона (таблица А10) и Спирмэна (таблица А11) при проверке нулевой гипотезы, что соответствующий коэффициент корреляции равен нулю (глава 26). Значимость достигается для данного объёма выборки при установленной величине *P*, если абсолютная величина (т.е. игнорируя её знак) значения выборочного коэффициента корреляции превышает табулированную величину. Например, если объём выборки равен 24 и *r* Пирсона равен 0,58, то 0,001 < *P* <0,01. Если размер выборки равен 7, а r_s Спирмэна = −0,63, тогда *P* >0,05.

Таблица 12 содержит числа 0–9, расположенные в случайном порядке.

Таблица А1. Стандартное нормальное распределение

z	Двустороннее значение *P*
0,0	1,000
0,1	0,920
0,2	0,841
0,3	0,764
0,4	0,689
0,5	0,617
0,6	0,549
0,7	0,484
0,8	0,424
0,9	0,368
1,0	0,317
1,1	0,271
1,2	0,230
1,3	0,194
1,4	0,162
1,5	0,134
1,6	0,110
1,7	0,089
1,8	0,072
1,9	0,057
2,0	0,046
2,1	0,036
2,2	0,028
2,3	0,021
2,4	0,016
2,5	0,012
2,6	0,009
2,7	0,007
2,8	0,005
2,9	0,004
3,0	0,003
3,1	0,002
3,2	0,001
3,3	0,001
3,4	0,001
3,5	0,000

Получено с использованием Microsoft Excel Version 5.0.

Таблица А2. *t*-распределение Стьюдента

df	Двустороннее значение *P*			
	0,10	0,05	0,01	0,001
1	6,314	12,706	63,656	636,58
2	2,920	4,303	9,925	31,600
3	2,353	3,182	5,841	12,924
4	2,132	2,776	4,604	8,610
5	2,015	2,571	4,032	6,869
6	1,943	2,447	3,707	5,959
7	1,895	2,365	3,499	5,408
8	1,860	2,306	3,355	5,041
9	1,833	2,262	3,250	4,781
10	1,812	2,228	3,169	4,587
11	1,796	2,201	3,106	4,437
12	1,782	2,179	3,055	4,318
13	1,771	2,160	3,012	4,221
14	1,761	2,145	2,977	4,140
15	1,753	2,131	2,947	4,073
16	1,746	2,120	2,921	4,015
17	1,740	2,110	2,898	3,965
18	1,734	2,101	2,878	3,922
19	1,729	2,093	2,861	3,883
20	1,725	2,086	2,845	3,850
21	1,721	2,080	2,831	3,819
22	1,717	2,074	2,819	3,792
23	1,714	2,069	2,807	3,768
24	1,711	2,064	2,797	3,745
25	1,708	2,060	2,787	3,725
26	1,706	2,056	2,779	3,707
27	1,703	2,052	2,771	3,689
28	1,701	2,048	2,763	3,674
29	1,699	2,045	2,756	3,660
30	1,697	2,042	2,750	3,646
40	1,684	2,021	2,704	3,551
50	1,676	2,009	2,678	3,496
100	1,660	1,984	2,626	3,390
200	1,653	1,972	2,601	3,340
5000	1,645	1,960	2,577	3,293

Получено с использованием Microsoft Excel Version 5.0

Таблица А3. Распределение Хи-квадрат Пирсона

df	Двустороннее значение *P*			
	0,10	0,05	0,01	0,001
1	2,706	3,841	6,635	10,827
2	4,605	5,991	9,210	13,815
3	6,251	7,815	11,345	16,266
4	7,779	9,488	13,277	18,466
5	9,236	11,070	15,086	20,515
6	10,645	12,592	16,812	22,457
7	12,017	14,067	18,475	24,321
8	13,362	15,507	20,090	26,124
9	14,684	16,919	21,666	27,877
10	15,987	18,307	23,209	29,588
11	17,275	19,675	24,725	31,264
12	18,549	21,026	26,217	32,909
13	19,812	22,362	27,688	34,527
14	21,064	23,685	29,141	36,124
15	22,307	24,996	30,578	37,698
16	23,542	26,296	32,000	39,252
17	24,769	27,587	33,409	40,791
18	25,989	28,869	34,805	42,312
19	27,204	30,144	36,191	43,819
20	28,412	31,410	37,566	45,314
21	29,615	32,671	38,932	46,796
22	30,813	33,924	40,289	48,268
23	32,007	35,172	41,638	49,728
24	33,196	36,415	42,980	51,179
25	34,382	37,652	44,314	52,619
26	35,563	38,885	45,642	54,051
27	36,741	40,113	46,963	55,475
28	37,916	41,337	48,278	56,892
29	39,087	42,557	49,588	58,301
30	40,256	43,773	50,892	59,702
40	51,805	55,758	63,691	73,403
50	63,167	67,505	76,154	86,660
60	74,397	79,082	88,379	99,608
70	85,527	90,531	100,43	112,32
80	96,578	101,88	112,33	124,84
90	107,57	113,15	124,12	137,21
100	118,50	124,34	135,81	149,45

Получено с использованием Microsoft Excel Version 5.0

Таблица А4. Стандартное нормальное распределение

	Двустороннее значение *P*				
	0,50	0,10	0,05	0,01	0,001
Relevant CI	50%	90%	95%	99%	99,9%
z (i.e. CI multiplier)	0,67	1,64	1,96	2,58	3,29

Получено с использованием Microsoft Excel Version 5.0

Таблица А6. Знаковый критерий

	k=число «положительных разностей» (см. пояснения)					
n′	0	1	2	3	4	5
4	0,125	0,624	1,000			
5	0,062	0,376	1,000			
6	0,032	0,218	0,688	1,000		
7	0,016	0,124	0,454	1,000		
8	0,008	0,070	0,290	0,726	1,000	
9	0,004	0,040	0,180	0,508	1,000	
10	0,001	0,022	0,110	0,344	0,754	1,000

Получено с использованием Microsoft Excel Version 5.0

Таблица А5. Таблица F-распределения

Степени свободы знаменателя	Двустороннее значение *P*	Одностороннее значение *P*	Степени свободы числителя												
			1	2	3	4	5	6	7	8	9	10	15	25	500
1	**0,05**	**0,25**	**647,8**	**799,5**	**864,2**	**899,6**	**921,8**	**937,1**	**948,2**	**956,6**	**963,3**	**968,6**	**984,9**	**998,1**	**1017,0**
1	0,10	0,05	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	245,9	249,3	254,1
2	**0,05**	**0,25**	**38,51**	**39,00**	**39,17**	**39,25**	**39,30**	**39,33**	**39,36**	**39,37**	**39,39**	**39,40**	**39,43**	**39,46**	**39,50**
2	0,10	0,05	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,43	19,46	19,49
3	**0,05**	**0,25**	**17,44**	**16,04**	**15,44**	**15,10**	**14,88**	**14,73**	**14,62**	**14,54**	**14,47**	**14,42**	**14,25**	**14,12**	**13,91**
3	0,10	0,05	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,63	8,53
4	**0,05**	**0,25**	**12,22**	**10,65**	**9,98**	**9,60**	**9,36**	**9,20**	**9,07**	**8,98**	**8,90**	**8,84**	**8,66**	**8,50**	**8,27**
4	0,10	0,05	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,77	5,64
5	**0,05**	**0,25**	**10,01**	**8,43**	**7,76**	**7,39**	**7,15**	**6,98**	**6,85**	**6,76**	**6,68**	**6,62**	**6,43**	**6,27**	**6,03**
5	0,10	0,05	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,52	4,37
6	**0,05**	**0,25**	**8,81**	**7,26**	**6,60**	**6,23**	**5,99**	**5,82**	**5,70**	**5,60**	**5,52**	**5,46**	**5,27**	**5,11**	**4,86**
6	0,10	0,05	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,83	3,68
7	**0,05**	**0,25**	**8,07**	**6,54**	**5,89**	**5,52**	**5,29**	**5,12**	**4,99**	**4,90**	**4,82**	**4,76**	**4,57**	**4,40**	**4,16**
7	0,10	0,05	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,40	3,24
8	**0,05**	**0,25**	**7,57**	**6,06**	**5,42**	**5,05**	**4,82**	**4,65**	**4,53**	**4,43**	**4,36**	**4,30**	**4,10**	**3,94**	**3,68**
8	0,10	0,05	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,11	2,94
9	**0,05**	**0,25**	**7,21**	**5,71**	**5,08**	**4,72**	**4,48**	**4,32**	**4,20**	**4,10**	**4,03**	**3,96**	**3,77**	**3,60**	**3,35**
9	0,10	0,5	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,89	2,72
10	**0,05**	**0,25**	**6,94**	**5,46**	**4,83**	**4,47**	**4,24**	**4,07**	**3,95**	**3,85**	**3,78**	**3,72**	**3,52**	**3,35**	**3,09**
10	0,10	0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,73	2,55
15	**0,05**	**0,25**	**6,20**	**4,77**	**4,15**	**3,80**	**3,58**	**3,41**	**3,29**	**3,20**	**3,12**	**3,06**	**2,86**	**2,69**	**2,41**
15	0,10	0,05	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,28	2,08
20	**0,05**	**0,25**	**5,87**	**4,46**	**3,86**	**3,51**	**3,29**	**3,13**	**3,01**	**2,91**	**2,84**	**2,77**	**2,57**	**2,40**	**2,10**
20	0,10	0,05	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,07	1,86
30	**0,05**	**0,25**	**5,57**	**4,18**	**3,59**	**3,25**	**3,03**	**2,87**	**2,75**	**2,65**	**2,57**	**2,51**	**2,31**	**2,12**	**1,81**
30	0,10	0,05	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,88	1,64
50	**0,05**	**0,25**	**5,34**	**3,97**	**3,39**	**3,05**	**2,83**	**2,67**	**2,55**	**2,46**	**2,38**	**2,32**	**2,11**	**1,92**	**1,57**
50	0,10	0,05	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,73	1,46
100	**0,05**	**0,25**	**5,18**	**3,83**	**3,25**	**2,92**	**2,70**	**2,54**	**2,42**	**2,32**	**2,24**	**2,18**	**1,97**	**1,77**	**1,38**
100	0,10	0,05	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,62	1,31
1000	**0,05**	**0,25**	**5,04**	**3,70**	**3,13**	**2,80**	**2,58**	**2,42**	**2,30**	**2,20**	**2,13**	**2,06**	**1,85**	**1,64**	**1,16**
1000	0,10	0,05	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,68	1,52	1,13

Получено с использованием Microsoft Excel Version 5.0

Таблица А7. Ранги доверительных интервалов для медианы

Объём выборки	Приблизительно		
	90% CI	95% CI	99% CI
6	1,6	1,6	–
7	1,7	1,7	–
8	2,7	1,8	–
9	2,8	2,8	1,9
10	2,9	2,9	1,10
11	3,9	2,10	1,11
12	3,10	3,10	2,11
13	4,10	3,11	2,12
14	4,11	3,12	2,13
15	4,12	4,12	3,13
16	5,12	4,13	3,14
17	5,13	4,14	3,15
18	6,13	5,14	4,15
19	6,14	5,15	4,16
20	6,15	6,15	4,17
21	7,15	6,16	5,17
22	7,16	6,17	5,18
23	8,16	7,17	5,19
24	8,17	7,18	6,19
25	8,18	8,18	6,20
26	9,18	8,19	6,21
27	9,19	8,20	7,21
28	10,19	9,20	7,22
29	10,20	9,21	8,22
30	11,20	10,21	8,23
31	11,21	10,22	8,24
32	11,22	10,23	9,24
33	12,22	11,23	9,25
34	12,23	11,24	9,26
35	12,23	12,24	10,26
36	13,24	12,25	10,27
37	14,24	13,25	11,27
38	14,25	13,26	11,28
39	14,26	13,27	11,29
40	15,26	14,27	12,29
41	15,27	14,28	12,30
42	16,27	15,28	13,30
43	16,28	15,29	13,31
44	17,28	15,30	13,32
45	17,29	16,30	14,32
46	17,30	16,31	14,33
47	18,30	17,31	15,33
48	18,31	17,32	15,34
49	19,31	18,32	15,35
50	19,32	18,33	16,35

Получено с использованием Microsoft Excel Version 5.0

Таблица А8. Критерий знаковых рангов Вилкоксона

	Двустороннее значение *P*		
n'	0,05	0,01	0,001
6	0–21	–	–
7	2–26	–	–
8	3–33	0–36	–
9	5–40	1–44	–
10	8–47	3–52	–
11	10–56	5–61	0–66
12	13–65	7–71	1–77
13	17–74	9–82	2–89
14	21–84	12–93	4–101
15	25–95	15–105	6–114
16	29–107	19–117	9–127
17	34–119	23–130	11–142
18	40–131	27–144	14–157
19	46–144	32–158	18–172
20	52–158	37–173	21–189
21	58–173	42–189	26–205
22	66–187	48–205	30–223
23	73–203	54–222	35–241
24	81–219	61–239	40–260
25	89–236	68–257	45–280

Адаптировано из Altman D.G. Practical Statistics for Medical Research. Copyright CRC Press, Bocu Raton, 1991.

Таблица А9(а). Критерий ранговых сумм Вилкоксона для двустороннего значения P=0,05

	n_s (число наблюдений в меньшей выборке)											
n_L	4	5	6	7	8	9	10	11	12	13	14	15
4	10–26	16–34	23–43	31–53	40–64	49–77	60–90	72–104	85–119	99–135	114–152	130–170
5	11–29	17–38	24–48	33–58	42–70	52–83	63–97	75–112	89–127	103–144	118–162	134–181
6	12–32	18–42	26–52	34–64	44–76	55–89	66–104	79–119	92–136	107–153	122–172	139–191
7	13–35	20–45	27–57	36–69	46–82	57–96	69–111	82–127	96–144	111–162	127–181	144–201
8	14–38	21–49	29–61	38–74	49–87	60–102	72–118	85–135	100–152	115–171	131–191	149–211
9	14–42	22–53	31–65	40–79	51–93	62–109	75–125	89–142	104–160	119–180	136–200	154–221
10	15–45	23–57	32–70	42–84	53–99	65–115	78–132	92–150	107–169	124–188	141–209	159–231
11	16–48	24–61	34–74	44–89	55–105	68–121	81–139	96–157	111–177	128–197	145–219	164–241
12	17–51	26–64	35–79	46–94	58–110	71–127	84–146	99–165	115–185	132–206	150–228	169–251
13	18–54	27–68	37–83	48–99	60–116	73–134	88–152	103–172	119–193	136–215	155–237	174–261
14	19–57	28–72	38–88	50–104	62–122	76–140	91–159	106–180	123–201	141–223	160–246	179–271
15	20–60	29–76	40–92	52–109	65–127	79–146	94–166	110–187	127–209	145–232	164–256	184–281

Таблица А9(b). Критерий ранговых сумм Вилкоксона для двустороннего значения P=0,01

	n_s (число наблюдений в меньшей выборке)											
n_L	4	5	6	7	8	9	10	11	12	13	14	15
4	–	–	21–45	28–56	37–67	46–80	57–93	68–108	81–123	94–140	109–157	125–175
5	–	15–40	22–50	29–62	38–74	48–87	59–101	71–116	84–132	98–149	112–168	128–187
6	10–34	16–44	23–55	31–67	40–80	50–94	61–109	73–125	87–141	101–159	116–178	132–198
7	10–38	16–49	24–60	32–73	42–86	52–101	64–116	76–133	90–150	104–169	120–188	136–209
8	11–48	17–53	25–65	34–78	43–93	54–108	66–124	79–141	93–159	108–178	123–199	140–220
9	11–45	18–57	26–70	35–84	45–99	56–115	68–132	82–149	96–168	111–188	127–209	144–231
10	12–48	19–61	27–75	37–89	47–105	58–122	71–139	84–158	99–177	115–197	131–219	149–241
11	12–52	20–65	28–80	38–95	49–111	61–128	73–147	87–166	102–186	118–207	135–229	153–252
12	13–55	21–69	30–84	40–100	51–117	63–135	76–154	90–174	105–195	122–216	139–239	157–263
13	13–59	22–73	31–89	41–106	53–123	65–142	79–161	93–182	109–203	125–226	143–249	162–273
14	14–62	22–78	32–94	43–111	54–130	67–149	81–169	96–190	112–212	129–235	147–259	166–284
15	15–65	23–82	33–99	44–117	56–136	69–156	84–176	99–198	115–221	133–244	151–269	171–294

Заимствовано из Diem K. Documents Geigy Scientific Tables. 7th edn, Blackwell Publishing. Oxford, 1970.

Таблица A10. Коэффициенты корреляции Пирсона

Объём выборки	Двустороннее значение *P*		
	0,05	**0,01**	**0,001**
5	0,878	0,959	0,991
6	0,811	0,917	0,974
7	0,755	0,875	0,951
8	0,707	0,834	0,925
9	0,666	0,798	0,898
10	0,632	0,765	0,872
11	0,602	0,735	0,847
12	0,576	0,708	0,823
13	0,553	0,684	0,801
14	0,532	0,661	0,780
15	0,514	0,641	0,760
16	0,497	0,623	0,742
17	0,482	0,606	0,725
18	0,468	0,590	0,708
19	0,456	0,575	0,693
20	0,444	0,561	0,679
21	0,433	0,549	0,665
22	0,423	0,537	0,652
23	0,413	0,526	0,640
24	0,404	0,515	0,629
25	0,396	0,505	0,618
26	0,388	0,496	0,607
27	0,381	0,487	0,597
28	0,374	0,479	0,588
29	0,367	0,471	0,579
30	0,361	0,463	0,570
35	0,334	0,430	0,532
40	0,312	0,403	0,501
45	0,294	0,380	0,474
50	0,279	0,361	0,451
55	0,266	0,345	0,432
60	0,254	0,330	0,414
70	0,235	0,306	0,385
80	0,220	0,286	0,361
90	0,207	0,270	0,341
100	0,217	0,283	0,357
150	0,160	0,210	0,266

Заимствовано из Diem K. Documents Geigy Scientific Tables. 7th edn, Blackwell Publishing. Oxford, 1970.

Таблица A11. Коэффициенты корреляции Спирмэна

Объём выборки	Двустороннее значение *P*		
	0,05	**0,01**	**0,001**
5	1,000		
6	0,886	1,000	
7	0,786	0,929	1,000
8	0,738	0,881	0,976
9	0,700	0,833	0,933
10	0,648	0,794	0,903

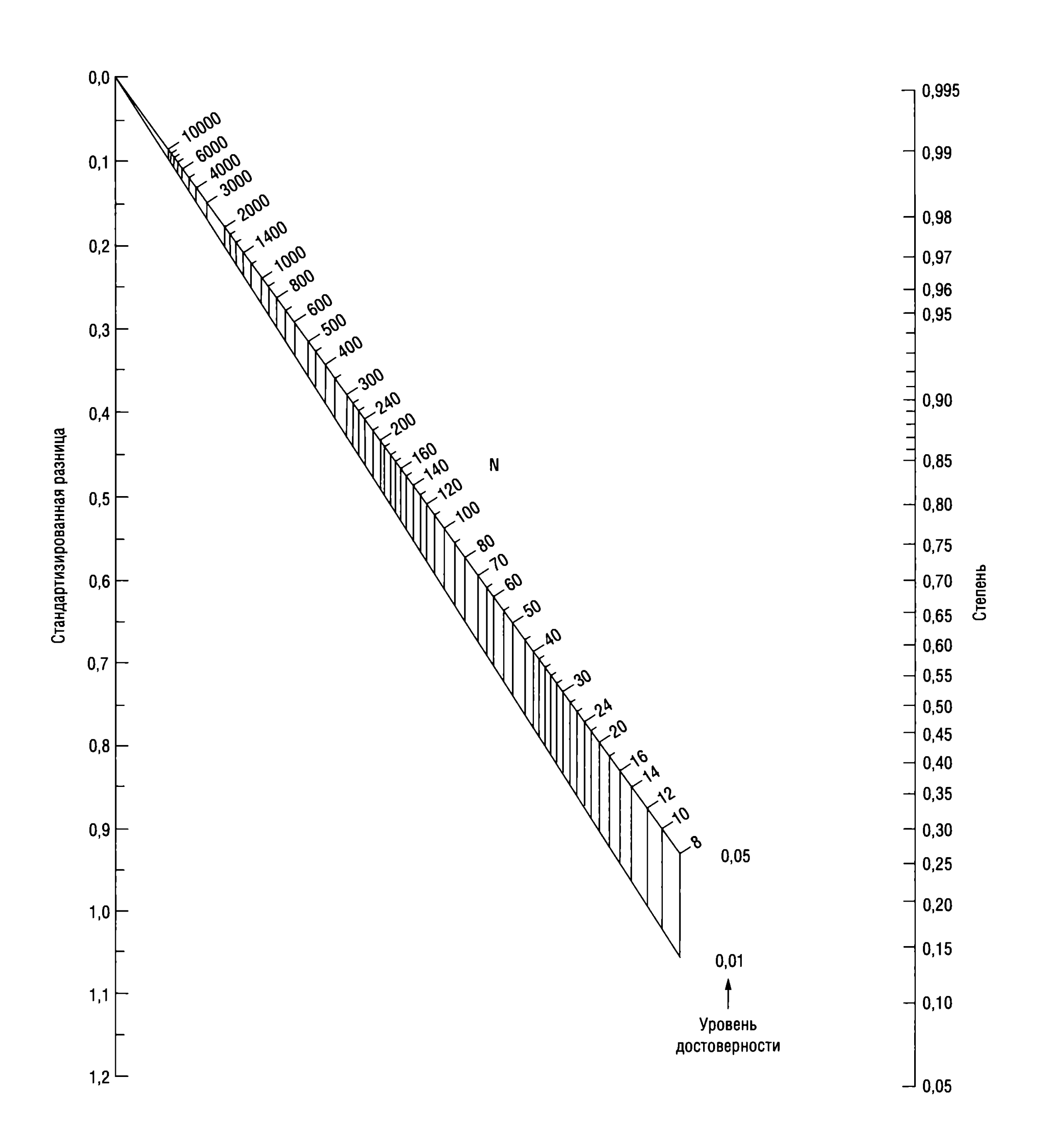
Адаптировано из Siegel S., Castellan N.J. Nonparametric Statistics for the Behavioural Sciences, 2nd edn, McGraw–Hill, New York, 1988, и использовано с разрешения издательства McGraw–Hill Companies.

Таблица A12. Таблица случайных чисел

3 4 8 1 4	6 8 0 2 0	2 8 9 9 8	5 1 6 8 7	4 0 0 8 8	3 5 4 5 8	2 4 7 0 8	0 1 8 1 5	5 3 7 7 6
9 9 1 0 6	5 0 8 9 9	0 7 3 9 4	9 1 0 7 1	2 2 4 1 1	6 1 6 4 3	6 4 4 3 5	6 2 5 5 2	6 4 3 1 6
4 7 1 8 5	3 1 7 8 2	4 8 8 9 4	6 8 7 9 0	5 1 8 5 2	3 6 9 1 8	0 5 7 3 7	9 0 6 5 3	6 1 1 2 3
8 1 3 5 4	5 7 2 9 6	3 9 3 2 9	5 2 2 6 3	4 3 1 9 4	5 1 6 2 4	4 2 4 2 9	6 1 3 6 7	4 1 2 0 7
8 3 4 6 7	8 5 6 2 2	9 5 7 7 8	0 5 3 4 7	0 0 4 4 5	5 1 3 3 4	2 9 4 4 5	9 9 1 7 6	3 0 0 9 1
2 7 9 2 4	3 4 1 6 7	5 7 0 6 0	5 7 5 3 5	3 2 2 7 8	1 6 9 4 9	0 4 9 6 0	0 4 1 1 6	9 1 4 6 7
5 8 3 1 9	8 8 1 6 4	9 4 1 3 0	0 7 7 4 3	1 6 9 1 7	1 5 6 8 1	9 3 5 7 2	9 9 7 5 3	4 9 1 1 7
4 9 7 3 2	6 6 7 0 2	7 2 4 2 5	9 9 1 1 7	4 9 2 9 8	8 7 2 6 5	1 4 1 9 5	8 3 3 9 1	1 9 7 9 4
6 9 5 9 4	2 6 7 4 9	6 8 7 4 3	3 9 1 3 9	4 4 4 9 5	1 1 9 4 4	1 2 9 7 0	5 6 5 2 3	6 2 4 1 1
3 0 0 7 4	9 7 5 1 7	9 7 4 5 0	5 4 2 5 1	5 1 7 7 7	2 1 0 7 3	0 3 9 0 9	2 6 5 1 9	3 9 5 7 8
8 1 1 4 7	5 7 5 0 8	9 3 4 7 9	8 7 8 2 6	2 8 9 6 5	7 4 4 7 4	9 7 4 6 8	8 0 1 4 9	1 7 8 3 4
7 4 6 8 9	2 8 9 3 3	5 9 8 1 9	9 3 0 5 2	6 1 3 2 5	8 3 1 4 5	4 4 6 8 4	7 2 9 5 8	9 1 8 2 4
1 4 8 0 2	2 5 9 8 2	4 8 0 2 4	1 5 4 6 1	3 7 5 7 0	4 4 6 8 5	4 7 3 8 6	0 9 5 0 4	7 7 8 3 1
6 8 5 0 1	3 4 1 9 4	8 5 3 5 5	3 8 4 1 1	4 6 5 5 9	4 1 6 9 4	9 9 6 7 8	8 8 2 6 8	8 6 6 7 4
4 8 7 3 4	9 2 6 7 1	8 5 2 5 2	8 5 9 8 5	3 4 2 2 8	9 1 2 8 9	5 6 3 3 1	1 4 6 8 3	3 6 4 9 3
8 4 1 0 2	8 1 6 9 9	9 7 3 5 2	5 4 5 0 9	9 3 1 9 6	5 1 2 0 4	4 3 3 5 1	1 1 8 1 8	4 1 1 7 9
2 8 4 3 2	3 2 8 7 3	8 3 8 3 4	0 9 8 6 2	1 2 7 2 0	6 4 5 6 9	4 2 2 1 8	2 6 7 2 6	8 0 8 6 6
9 1 4 5 8	8 2 5 2 4	7 5 5 2 3	0 1 2 7 6	1 9 5 9 1	4 7 4 7 3	9 0 2 5 1	9 9 1 0 3	7 2 9 4 7
4 5 4 3 5	3 0 3 8 9	6 9 7 3 2	8 1 9 6 2	3 0 2 4 3	9 6 1 9 9	3 3 5 4 6	3 9 6 7 2	8 3 7 6 0
2 3 5 5 7	7 8 4 3 7	4 4 9 5 7	9 8 7 2 8	6 5 6 7 4	3 4 7 0 1	8 3 3 9 8	5 4 1 0 2	6 5 8 4 5
3 0 3 9 5	9 1 8 5 0	5 2 0 0 4	0 4 8 4 4	2 8 8 4 8	1 9 7 2 8	9 6 5 7 1	1 3 3 1 7	7 0 8 5 9
6 9 9 9 1	1 2 7 5 5	9 7 9 1 6	5 7 6 3 9	4 3 4 4 5	9 0 4 6 3	8 5 5 5 6	3 5 4 6 9	1 9 7 4 9
3 2 9 8 0	4 3 6 0 8	2 0 5 9 2	7 2 5 2 7	6 3 5 8 3	4 6 4 4 3	5 3 9 2 9	8 7 2 1 9	5 5 1 9 8
5 9 7 7 6	3 7 0 3 5	5 3 7 6 5	5 5 1 9 6	6 8 6 5 9	7 1 4 2 9	2 5 2 2 5	9 1 9 4 2	5 1 1 3 2
7 3 7 1 4	7 9 8 6 8	2 3 8 8 0	9 2 2 5 4	7 2 9 8 4	0 7 7 9 2	8 1 3 0 6	2 4 2 7 7	8 2 3 6 6
6 1 5 4 7	1 6 5 7 5	6 8 5 2 0	5 9 8 6 9	6 7 2 9 9	7 3 5 6 5	7 7 3 1 6	9 6 6 8 2	1 8 0 3 1
8 7 7 3 7	0 1 0 5 8	7 6 0 1 2	7 6 2 4 7	7 5 6 1 6	5 1 3 3 5	7 0 3 6 4	7 8 9 4 2	4 0 5 6 4
9 8 6 6 9	0 8 3 3 4	4 0 5 2 0	7 8 3 8 9	5 6 4 9 8	7 4 3 3 6	0 2 4 3 4	4 8 5 9 9	6 7 5 7 9
8 1 5 3 5	4 6 6 9 0	9 2 8 1 4	4 4 4 5 6	2 9 2 2 7	4 8 1 2 2	3 0 5 2 2	1 3 8 5 2	4 8 4 3 6
0 5 9 7 5	4 7 1 1 0	3 2 7 3 3	4 6 9 2 9	9 8 2 6 1	5 2 1 9 3	8 3 2 1 5	5 3 1 9 2	8 3 1 0 9

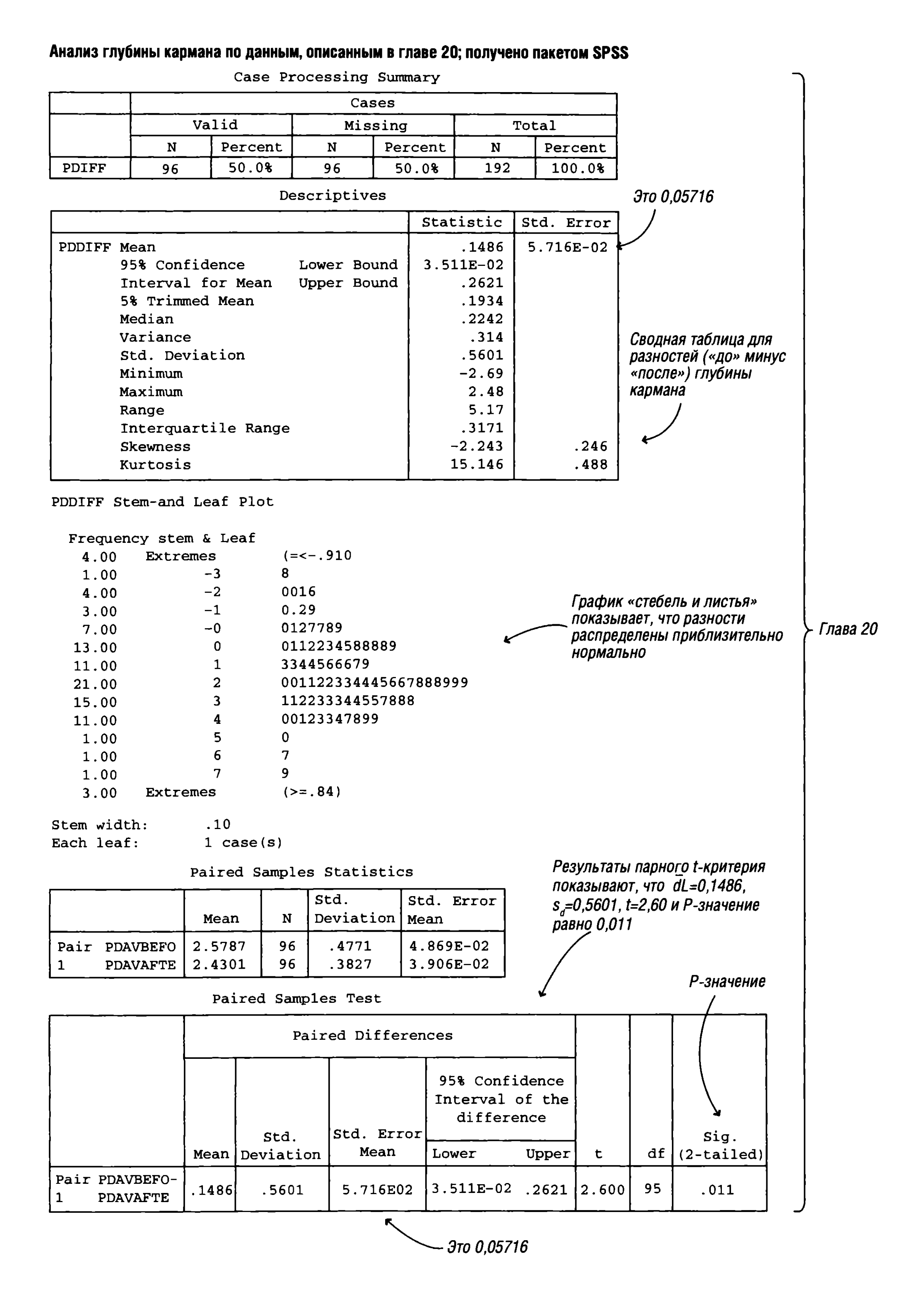
Получено с использованием Microsoft Excel Version 5.0.

Приложение Б. Номограмма Альтмана для определения объёма выборки (глава 36)



Заимствовано из: Altman D.G. How large a sample? // Statistics in Practice (eds. S.M. Gore & D.G. Altman). BMA, London. Copyright, 1982. С разрешения издательства B.M.J. Publishing Group.

Приложение В. Типичные компьютерные листинги результатов анализа



Анализ глубины кармана по данным, описанным в главе 20; получено пакетом SPSS

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PDIFF	96	50.0%	96	50.0%	192	100.0%

Descriptives

			Statistic	Std. Error
PDDIFF	Mean		.1486	5.716E-02
	95% Confidence Interval for Mean	Lower Bound	3.511E-02	
		Upper Bound	.2621	
	5% Trimmed Mean		.1934	
	Median		.2242	
	Variance		.314	
	Std. Deviation		.5601	
	Minimum		-2.69	
	Maximum		2.48	
	Range		5.17	
	Interquartile Range		.3171	
	Skewness		-2.243	.246
	Kurtosis		15.146	.488

PDDIFF Stem-and Leaf Plot

```
 Frequency stem & Leaf
   4.00   Extremes      (=<-.910
   1.00         -3      8
   4.00         -2      0016
   3.00         -1      0.29
   7.00         -0      0127789
  13.00          0      0112234588889
  11.00          1      3344566679
  21.00          2      001122334445667888999
  15.00          3      112233344557888
  11.00          4      00123347899
   1.00          5      0
   1.00          6      7
   1.00          7      9
   3.00   Extremes      (>=.84)

Stem width:      .10
Each leaf:       1 case(s)
```

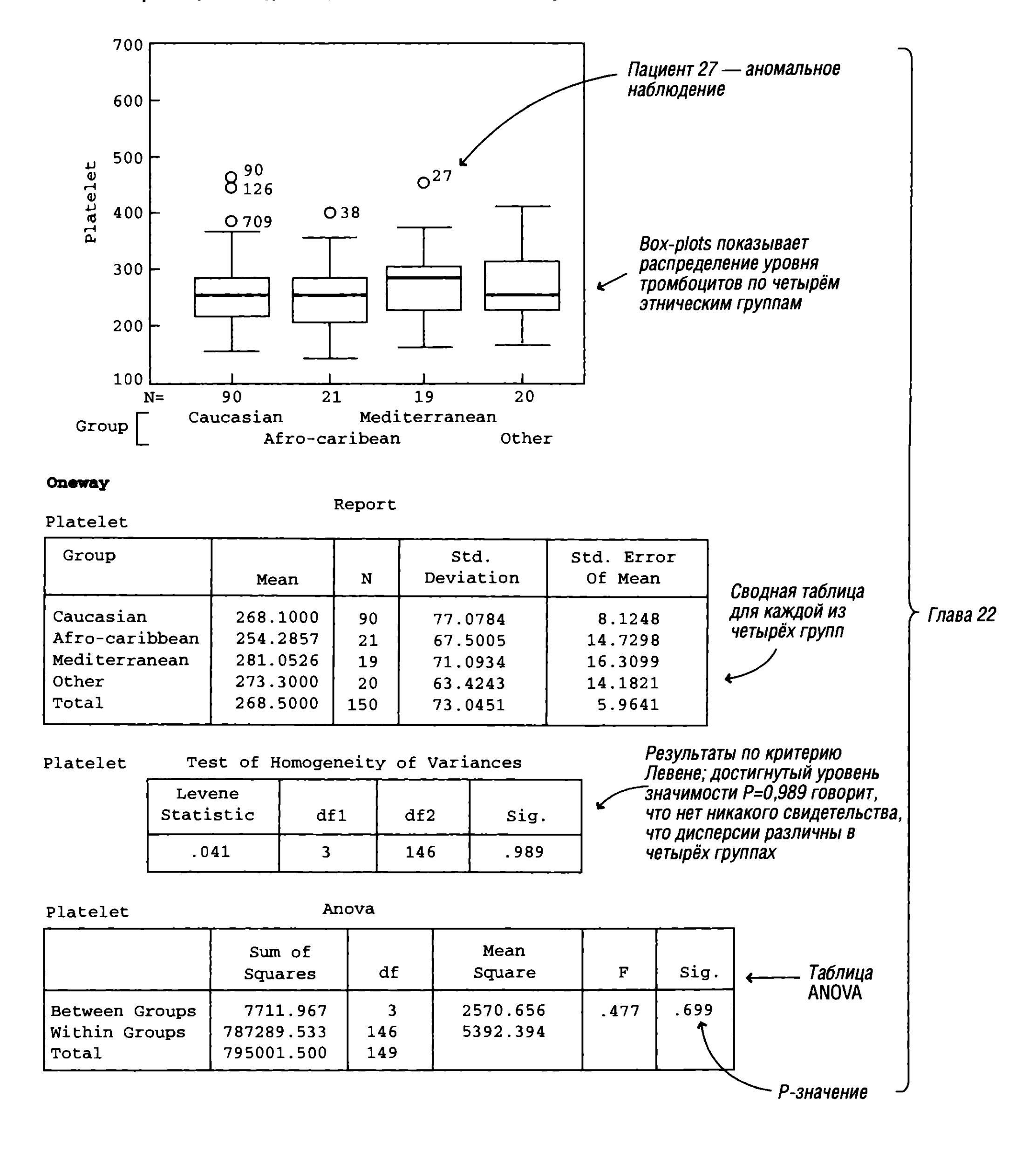
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PDAVBEFO	2.5787	96	.4771	4.869E-02
	PDAVAFTE	2.4301	96	.3827	3.906E-02

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the difference				
				Lower	Upper			
Pair 1 PDAVBEFO-PDAVAFTE	.1486	.5601	5.716E02	3.511E-02	.2621	2.600	95	.011

Анализ тромбоцитов по данным, описанным в главе 22; получено пакетом SPSS



Oneway

Report

Platelet

Group	Mean	N	Std. Deviation	Std. Error Of Mean
Caucasian	268.1000	90	77.0784	8.1248
Afro-caribbean	254.2857	21	67.5005	14.7298
Mediterranean	281.0526	19	71.0934	16.3099
Other	273.3000	20	63.4243	14.1821
Total	268.5000	150	73.0451	5.9641

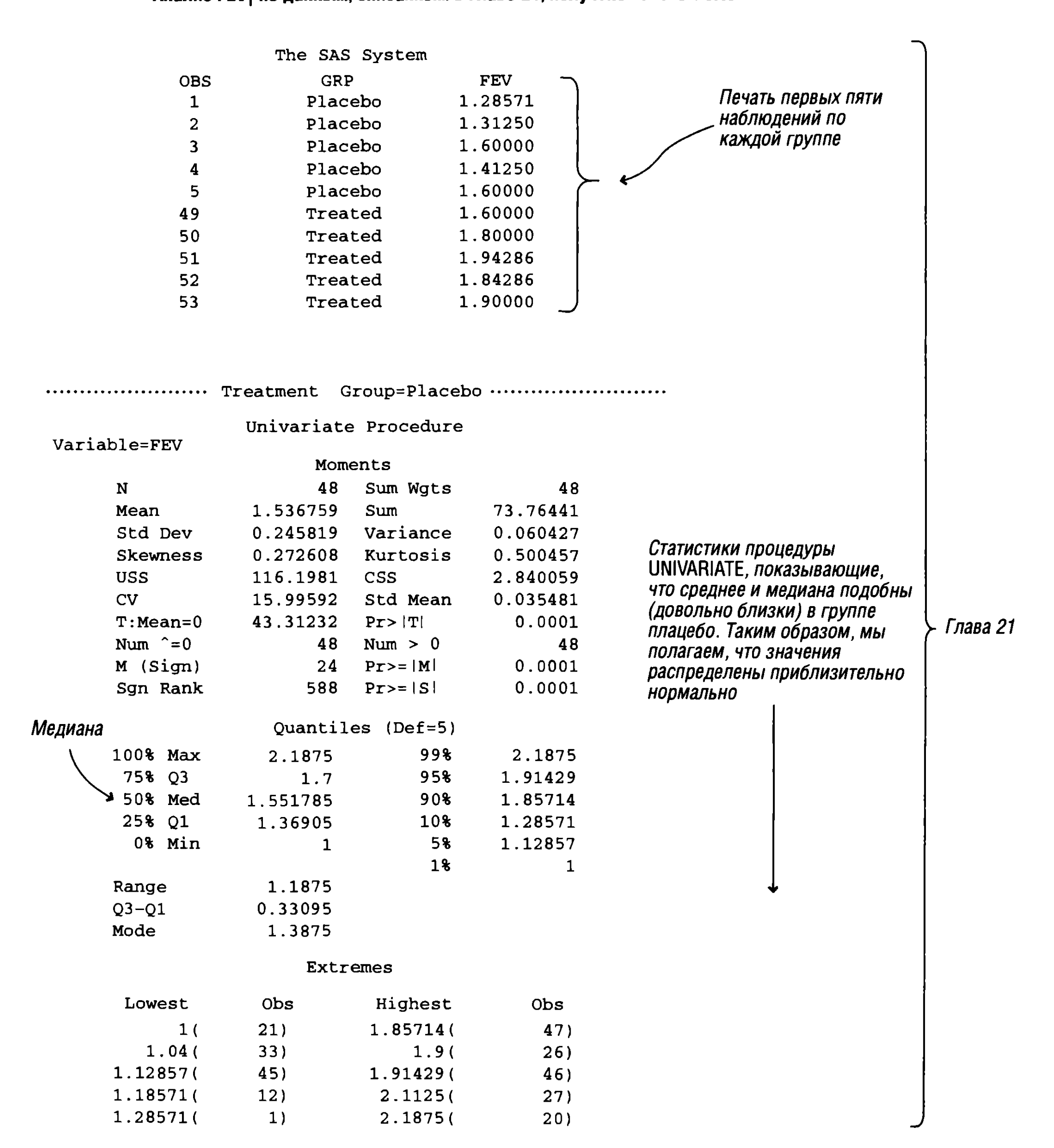
Platelet Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
.041	3	146	.989

Platelet Anova

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7711.967	3	2570.656	.477	.699
Within Groups	787289.533	146	5392.394		
Total	795001.500	149			

Анализ FEV_1 по данным, описанным в главе 21; получено пакетом SAS



```
                 The SAS System
        OBS         GRP           FEV
         1         Placebo       1.28571
         2         Placebo       1.31250
         3         Placebo       1.60000
         4         Placebo       1.41250
         5         Placebo       1.60000
        49         Treated       1.60000
        50         Treated       1.80000
        51         Treated       1.94286
        52         Treated       1.84286
        53         Treated       1.90000


....................... Treatment  Group=Placebo .........................

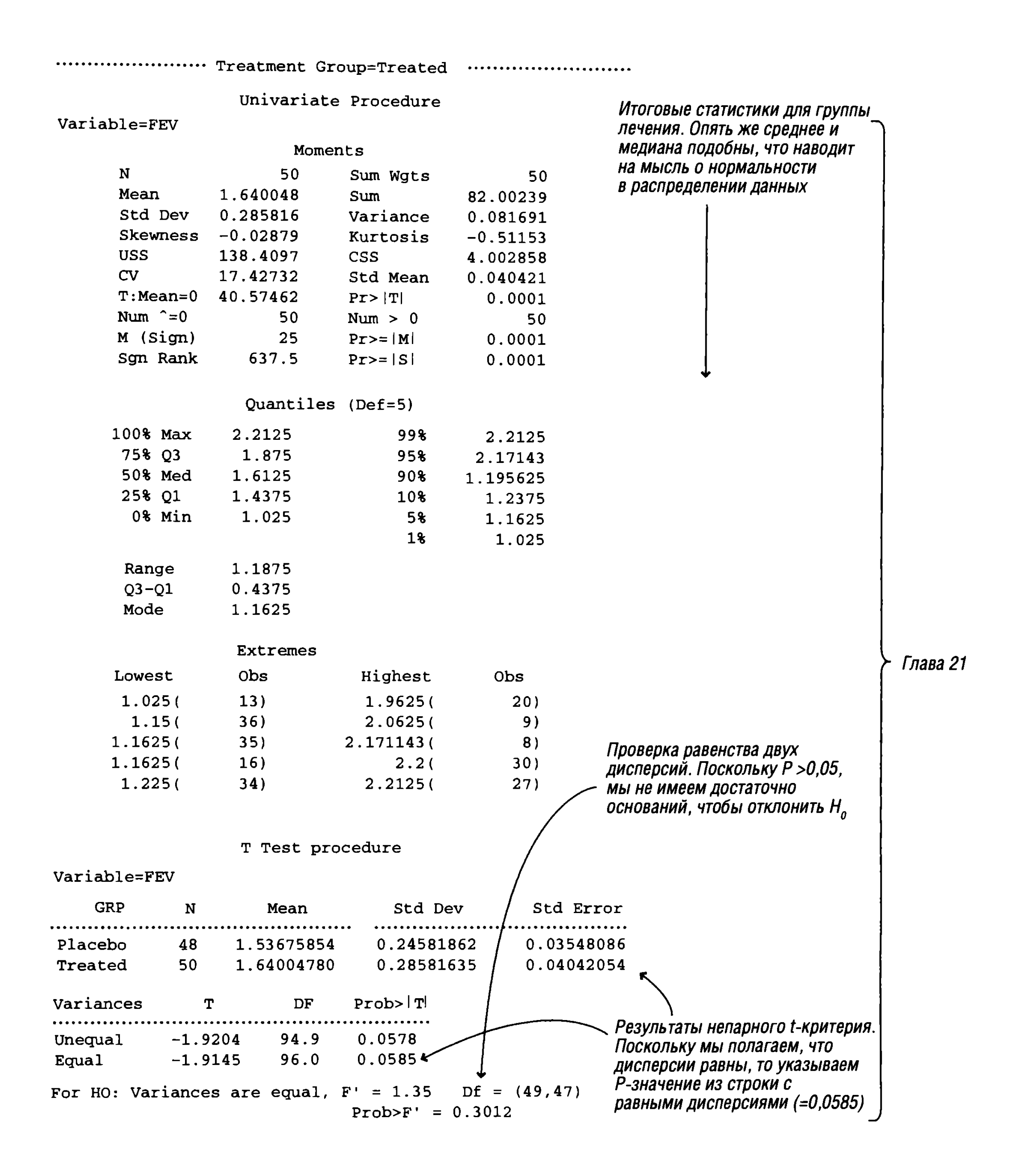
               Univariate Procedure
Variable=FEV
                      Moments
      N                 48    Sum Wgts           48
      Mean        1.536759    Sum          73.76441
      Std Dev     0.245819    Variance     0.060427
      Skewness    0.272608    Kurtosis     0.500457
      USS         116.1981    CSS          2.840059
      CV          15.99592    Std Mean     0.035481
      T:Mean=0    43.31232    Pr>|T|         0.0001
      Num ^=0           48    Num > 0            48
      M (Sign)          24    Pr>=|M|        0.0001
      Sgn Rank         588    Pr>=|S|        0.0001

                  Quantiles (Def=5)
      100% Max      2.1875        99%      2.1875
       75% Q3          1.7        95%     1.91429
       50% Med    1.551785        90%     1.85714
       25% Q1      1.36905        10%     1.28571
        0% Min           1         5%     1.12857
                                   1%           1
      Range         1.1875
      Q3-Q1        0.33095
      Mode          1.3875

                     Extremes

       Lowest      Obs       Highest        Obs
            1(     21)      1.85714(        47)
         1.04(     33)          1.9(        26)
      1.12857(     45)      1.91429(        46)
      1.18571(     12)       2.1125(        27)
      1.28571(      1)       2.1875(        20)
```


continued



Treatment Group=Treated
Univariate Procedure
Variable=FEV
Moments
N 50 Sum Wgts 50
Mean 1.640048 Sum 82.00239
Std Dev 0.285816 Variance 0.081691
Skewness -0.02879 Kurtosis -0.51153
USS 138.4097 CSS 4.002858
CV 17.42732 Std Mean 0.040421
T:Mean=0 40.57462 Pr>|T| 0.0001
Num ^=0 50 Num > 0 50
M (Sign) 25 Pr>=|M| 0.0001
Sgn Rank 637.5 Pr>=|S| 0.0001
Quantiles (Def=5)
100% Max 2.2125 99% 2.2125
75% Q3 1.875 95% 2.17143
50% Med 1.6125 90% 1.195625
25% Q1 1.4375 10% 1.2375
0% Min 1.025 5% 1.1625
1% 1.025
Range 1.1875
Q3-Q1 0.4375
Mode 1.1625
Extremes
Lowest Obs Highest Obs
1.025(13) 1.9625(20)
1.15(36) 2.0625(9)
1.1625(35) 2.171143(8)
1.1625(16) 2.2(30)
1.225(34) 2.2125(27)
T Test procedure
Variable=FEV
GRP N Mean Std Dev Std Error
Placebo 48 1.53675854 0.24581862 0.03548086
Treated 50 1.64004780 0.28581635 0.04042054
Variances T DF Prob>|T|
Unequal -1.9204 94.9 0.0578
Equal -1.9145 96.0 0.0585
For HO: Variances are equal, F' = 1.35 Df = (49,47) Prob>F' = 0.3012
Итоговые статистики для группы лечения. Опять же среднее и медиана подобны, что наводит на мысль о нормальности в распределении данных
Глава 21
Проверка равенства двух дисперсий. Поскольку P >0,05, мы не имеем достаточно оснований, чтобы отклонить H0
Результаты непарного t-критерия. Поскольку мы полагаем, что дисперсии равны, то указываем P-значение из строки с равными дисперсиями (=0,0585)

Анализ антропометрических данных, описанных в главах 26, 28 и 29; получено пакетом SAS

OBS	SBP	Height	Weight	Sex
1	91.00	119.7	20.0	0
2	122.50	124.6	42.5	0
3	109.50	111.3	19.8	0
4	100.50	110.3	18.9	0
5	99.00	112.5	19.0	0
6	103.50	115.1	19.3	0
7	101.00	116.3	19.6	0
8	103.00	111.1	17.1	1
9	106.50	117.2	20.7	1
10	102.50	113.2	22.1	1

Распечатка из данных по первым 10 детям

Correlation Analysis

4 'VAR' Variables: SBP Height Weight Age

Simple Statistics

Variable	N	Mean	Std Dev	Sum
SBP	100	104.414700	9.430933	10441
Height	100	120.054000	6.439986	12005
Weight	100	22.826000	4.223303	2282.600000
Age	100	6.696900	0.731717	669.690000

Simple Statistics

Variable	Minimum	Maximum
SBP	81.500000	128.850000
Height	107.1000000	136.800000
Weight	15.900000	42.500000
Age	5.130000	8.840000

Сводные статистики по каждой переменной

Pearson Correlation Coefficients/Prob> |R| under Ho:Rho=0
/N=100

	SBP	Height	Weight	Age
SBP	1.00000	0.33066	0.51774	0.16373
	0.0	0.0008	0.0001	0.1036
Height	0.33066	1.00000	0.69151	0.64486
	0.0008	0.0	0.0001	0.0001
Weight	0.51774	0.69151	1.00000	0.38935
	0.0001	0.0001	0.0	0.0001
Age	0.16373	0.64486	0.38935	1.00000
	0.1036	0.0001	0.0001	0.0

Коэффициент корреляции Пирсона между систолическим артериальным давлением (SBP) и возрастом (Age)

Соответствующее Р-значение

Spearman Correlation Coefficients/Prob> |R| under Ho:Rho=0
/N=100

	SBP	Height	Weight	Age
SBP	1.00000	0.31519	0.45453	0.14778
	0.0	0.0014	0.0001	0.1423
Height	0.31519	1.00000	0.82298	0.61491
	0.0014	0.0	0.0001	0.0001
Weight	0.45453	0.82298	1.00000	0.51260
	0.0001	0.0001	0.0	0.0001
Age	0.14778	0.61491	0.51260	1.00000
	0.1423	0.0001	0.0001	0.0

Коэффициент корреляции Спирмена между ростом и возрастом

Р-значение

Глава 26

Model:MODEL1
Dependent Variable:SBP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	962.71441	962.71441	12.030	0.0008
Error	98	7842.59208	80.02645		
C Total	99	8805.30649			

ANOVA-таблица

Root MSE	8.94575	R-square	0.1093
Dep Mean	104.41470	Adj R-sq	0.1002
C.V.	8.56752		

Parameter Estimates

Свободный член, пересечение

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
Intercep	1	46.281684	16.78450788	2.757
Height	1	0.484224	0.13960927	3.468

Коэффициент регрессии

Variable	DF	Prob>|T|
Intercep	1	0.0070
Height	1	0.0008

Результаты парной линейной регрессии систолического артериального давления (SBP) от роста Глава 28

Model:MODEL1
Dependent Variable:SBP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	2804.04514	934.68171	14.952	0.0001
Error	96	6001.26135	62.51314		
C Total	99	8805.30649			

Root MSE	7.90653	R-square	0.3184
Dep Mean	104.41470	Adj R-sq	0.2972
C.V.	7.57223		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0
Intercep	1	79.439541	17.11822110	4.641
Height	1	-0.031023	0.17170250	-0.181
Weight	1	1.179495	0.26139400	4.512
Sex	1	4.229540	1.61054848	2.626

Оценки отдельных регрессионных коэффициентов

Variable	DF	Prob>|T|
Intercep	1	0.0001
Height	1	0.8570
Weight	1	0.0001
Sex	1	0.0101

Результаты множественной линейной регрессии систолического артериального давления (SBP) от роста, веса и пола Глава 29

Анализ данных по вирусу герпеса человека HHV-8, описанному в главах 23, 24 и 30; получено пакетом STATA

```
.List hhv8 gonnorrho syphilis hsv2 hiv age in 1/10

        hhv8       gonorrho    syphilis   hsv2   hiv    age
 1.  negative      history     0          0      0      28
 2.  negative      history     0          0      0      40
 3.  negative      history     0          0      0      26
 4.  negative      history     0          1      0      42
 5.  positive      history     0          0      0      30
 6.  negative      nohistory   0          0      0      33
 7.  negative      history     0          1      0      27
 8.  positive      history     0          0      0      32
 9.  negative      history     1          0      0      35
10.  positive      history     0          0      0      35
```

Распечатка из данных по первым 10 мужчинам

```
   Tabulate gonorrho hhv8, chi row col

                    | hhv8          |              |
        gonorrhoe   | negative      |  positive    |     Total
       -------------+---------------+--------------+-------------
          History   |     192       |      36      |     (228)
                    |   84.21       |   15.79      |    100.00
                    |   86.88       |   72.00      |     84.13
       -------------+---------------+--------------+-------------
       No history   |      29       |     (14)     |      43
                    |   67.44       |   32.56      |    100.00
                    |   13.12       |   28.00      |     15.87
       -------------+---------------+--------------+-------------
            Total   |     221       |     (50)     |     (271)
                    |   81.55       |   18.45      |    100.00
                    |  100.00       |  100.00      |    100.00

                Pearson chi2(1) = 6.7609   Pr = 0.009
```

Таблица сопряжённости

Строка %

Колонка %

Частная сумма по строке

Наблюдаемые (фактические) частоты

Частная сумма по колонке (столбцу)

Общая сумма

Глава 24

```
. Logit hhv8 gonorrho syphilis hsv2 hiv age, or tab

Interation 0:   Log Likelihood = -122.86506
Interation 1:   Log Likelihood = -111.87072
Interation 2:   Log Likelihood = -110.58712
Interation 3:   Log Likelihood = -110.56596
Interation 4:   Log Likelihood = -110.56595

                                              Number of obs =    260
Logit Estimates                               chi2 (5)      =  24.60
                                              Prob > chi2   = 0.0002
Logit Likelihood = -110.56595                 Pseudo R2     = 0.1001

---------+------------------------------------------------------------------
    hhv8 |    Coef.     Std. Err.      z         P>|z|    [95% Conf. Interval]
---------+------------------------------------------------------------------
gonorrho | .5093263    .4363219      1.167     0.243     -.345849    1.364502
syphilis | 1.192442    .7110707      1.677     0.094     -.201231    2.586115
    hsv2 | .7910041    .3871114      2.043     0.041     .0322798    1.549728
     hiv | 1.635669    .6028147      2.713     0.007     .4541736    2.817164
     age | .0061609    .0204152      0.302     0.763    -.0338521     .046174
constant |-2.224164    .6511603     -3.416     0.001    -3.500415   -.9479135
---------+------------------------------------------------------------------

---------+------------------------------------------------------------------
    hhv8 |Odds Ratio   Std. Err.      z        P>|z|    [95% Conf. Interval]
---------+------------------------------------------------------------------
gonorrho |  1.66417    .7261137     1.167     0.243     .7076193    3.913772
syphilis | 3.295118    2.343062     1.677     0.094     .8177235    13.27808
    hsv2 |  2.20561    .8538167     2.043     0.041     1.032806    4.710191
     hiv | 5.132889    3.094181     2.713     0.007     1.574871    16.72934
     age |  1.00618    .0205413     0.302     0.763     .9667145    1.047257
---------+------------------------------------------------------------------

      Comparison of outcomes and probabilites

      Outcome |  Pr < .5      Pr > = .5 |     Total
     ---------+-------------------------+-------------
      Failure |   208               5   |      213
      Success |    38               9   |       47
     ---------+-------------------------+-------------
        Total |   246              14   |      260
```

Хи-квадрат для ковариат и его Р-значение

Р-значение

Доверительный интервал для отношения шансов

Предсказание исхода <0,5=0 (нет), ≥0,5=1 (да)

Наблюдаемый (фактический) исход Неуспех =0 (нет) Успех =1 (да)

Классификационная таблица

Результаты из множественной логистической регрессии Глава 30

Анализ данных по вирусологическим неудачам, описанным в главах 31-33, проведённых пакетом SAS

Время начала первичного отклика (<1 года =1, 1–2 года =2, >2 лет =3) → PERIOD
Вирусологическая неудача (нет=0, да=1) → EVENT
Продолжительность наблюдения, дни → PDAYS
Женщина=0, мужчина=1 → SEX
Статус лечения (предварительное общепринятое лечение =0, без предварительного лечения =1) → TRTSTATUS

Печать данных для первых 10 пациентов (по каждому пациенту — одна строка данных для каждого периода времени)

OBS	PATIENT	PERIOD	EVENT	PDAYS	SEX	BASECD8	TRTSTATUS
1	1	1	0	365.25	0	665	1
2	1	2	0	48.75	0	665	1
3	2	1	0	365.25	1	2053	1
4	2	2	0	365.25	1	2053	1
5	2	3	0	592.50	1	2053	1
7	4	1	0	30.00	0	327	1
8	5	1	0	365.25	1	931	1
9	5	2	0	365.25	1	931	1
10	5	3	0	732.50	1	931	1
13	6	1	0	166.00	1	1754	1
14	7	1	0	84.00	1	665	1
15	8	1	0	365.25	1	297	1
16	8	2	0	152.75	1	297	1
17	9	1	0	142.00	1	455	1
18	10	1	0	230.00	0	736	1

Модель 1, включающая две искусственные переменных для времени начала первичного отклика Глава 32

```
The GENMOD Procedure
Model Information
Data Set              WORK.APPENDIX_POISSON
Distribution             Poisson
Link Function              Log
Dependent Variable         EVENT
Offset Variable            LTIME = Log (PDAYS)
Observations Used           988

Criteria For Assessing Goodness Of Fit
Criterion               DF        Value      Value/DF

Deviance               984     393.1203       0.3995
Scaled Deviance        984     393.1203       0.3995
Pearson Chi-Square     984    7574.2725       7.6974
Scaled Pearson X2      984    7574.2725       7.6974
Log Likelihood                -257.5601

Analysis Of Parameter Estimates

                      Standard     Wald 95%            Chi-
Parameter  DF Estimate   Error   Confidence Limits    Square  Pr > ChiSq

Intercept   1  -1.1698  0.3228  -1.8024   -0.5372     13.14      0.0003
TRTSTATUS   1  -0.6096  0.2583  -1.1159   -0.1033      5.57      0.0183
BASECD8_100 1  -0.0587  0.0268  -0.1112   -0.0063      4.82      0.0281
SEX         1  -0.4923  0.2660  -1.0136    0.0290      3.43      0.0642
Scale       0   1.0000  0.0000   1.0000    1.0000

LR Statistics For Type 3 Analysis
                          Chi-
Source          DF      Square    Pr > ChiSq

TRTSTATUS        1        5.40        0.0201
BASECD8_100      1        5.46        0.0194
SEX              1        3.27        0.0707
```

Исходный уровень CD8, делённый на 100 → BASECD8_100

Статистика критерия Вальда → Chi-Square

Масштабный параметр, используемый для пуассоновской экстрадисперсии → Scale

Р-значения достигнутого уровня значимости для каждой переменной в модели → Pr > ChiSq

```
                     Model Information
           Data Set                 WORK.APPENDIX_POISSON
           Distribution                 Poisson
           Link Function                    Log
           Dependent Variable             EVENT
           Offset Variable                LTIME
           Observations Used                988

                  Class Level Information

           Class          Levels     Values
           PERIOD3             3     1 2 3

              Criteria For Assessing Goodness Of Fit
  Criterion                  DF          Value        Value/DF

  Deviance                  982       387.5904          0.3947
  Scaled Deviance           982       387.5904          0.3947
  Pearson Chi-Square        982      5890.6342          5.9986
  Scaled Pearson X2         982      5890.6342          5.9986
  Log Likelihood                     -254.7952

               Analysis Of Parameter Estimates

                            Standard      Wald 95%           Chi-
Parameter       DF  Estimate   Error   Confidence Limits    Square  Pr > ChiSq

Intercept        1   -1.2855  0.3400   -1.9518   -0.6192    14.30      0.0002
TRTSTATUS        1   -0.5871  0.2587   -1.0942   -0.0800     5.15      0.0233
BASECD8_100      1   -0.0558  0.0267   -0.1083   -0.0034     4.36      0.0369
SEX              1   -0.4868  0.2664   -1.0089    0.0353     3.34      0.0676
PERIOD      1    0    0.0000  0.0000    0.0000    0.0000      .          .
PERIOD      2    1    0.4256  0.2702   -0.1039    0.9552     2.48      0.1152
PERIOD      3    1   -0.5835  0.4825   -1.5292    0.3622     1.46      0.2265
Scale            0    1.0000  0.0000    1.0000    1.0000

              LR Statistics For Type 3 Analysis
                                      Chi-
           Source               DF   Square    Pr > ChiSq

           TRTSTATUS             1     5.00        0.0253
           BASECD8_100           1     4.91        0.0267
           SEX                   1     3.19        0.0742
           PERIOD                2     5.53        0.0630
```

СМП или отклонение даёт P >0,99 для оценки согласия подгонки (→ Deviance)

Степень свободы (→ DF)

Это значительно меньше 1, свидетельствует о недостаточной дисперсии (→ 0.3947)

Оценки параметров модели, показанные в таблице 31-1 — Относительные интенсивности, полученные как антилогарифмы этих интенсивностей (→ Estimate)

Доверительный интервал для коэффициентов модели (→ Confidence Limits)

Нули в этой строке говорят о том, что этот период 1 есть ссылочная градация (→ PERIOD 1)

Тест-статистика = разнице в отклонениях 2-х моделей =393,1203 – 387,5904 (→ 5.53)

P-значение для теста различия в отклонениях из моделей вместе и без индикаторных переменных для времени начала первичного отклика (→ 0.0630)

Степени свободы = разность в числе параметров в моделях 1 и 2 (→ 2)

Модель 2, включающая две индикаторные переменные для времени начала первичного отклика и CD8, учитываемую как числовая переменная Главы 31 и 32

```
                     Model Information
           Data Set                 WORK.APPENDIX_POISSON
           Distribution                 Poisson
           Link Function                    Log
           Dependent Variable             EVENT
           Offset Variable                LTIME
           Observations Used                988


                  Class Level Information
           Class          Levels     Values
           PERIOD              3      1 2 3

              Criteria For Assessing Goodness Of Fit

  Criterion                  DF          Value        Value/DF

  Deviance                  983       392.5001          0.3993
  Scaled Deviance           983       392.5001          0.3993
  Pearson Chi-Square        983      5580.2152          5.6767
  Scaled Pearson X2         983      5580.2152          5.6767
  Log Likelihood                     -257.2501

               Analysis Of Parameter Estimates

                            Standard      Wald 95%           Chi-
Parameter       DF  Estimate   Error   Confidence Limits    Square  Pr > ChiSq

Intercept        1   -1.7549  0.2713   -2.2866   -1.2232    41.85      <.0001
TRTSTATUS        1   -0.6290  0.2577   -1.1340   -0.1240     5.96      0.0146
SEX              1   -0.5444  0.2649   -1.0637   -0.0252     4.22      0.0399
PERIOD      1    0    0.0000  0.0000    0.0000    0.0000      .          .
PERIOD      2    1    0.4191  0.2701   -0.1103    0.9485     2.41      0.1207
PERIOD      3    1   -0.6481  0.4814   -1.5918    0.2955     1.81      0.1782
Scale            0    1.0000  0.0000    1.0000    1.0000

              LR Statistics For Type 3 Analysis
                                      Chi-
           Source               DF   Square    Pr > ChiSq

           TRTSTATUS             1     5.77        0.0163
           SEX                   1     4.00        0.0455
           PERIOD                2     6.08        0.0478
```

Модель, включающая исходный уровень CD8 Глава 33

Model Information

Data Set	WORK.APPENDIX_POISSON
Distribution	Poisson
Link Function	Log
Dependent Variable	EVENT
Offset Variable	LTIME
Observations Used	988

Class Level Information

Class	Levels	Values
PERIOD	3	1 2 3
CD8GRP	5	1 2 3 4 5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	979	387.1458	0.3955
Scaled Deviance	979	387.1458	0.3955
Pearson Chi-Square	979	5852.1596	5.9777
Scaled Pearson X2	979	5852.1596	5.9777
Log Likelihood		-254.5729	

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-1.2451	0.6116	-2.4439	-0.0463	4.14	0.0418
TRTSTATUS		1	-0.5580	0.2600	-1.0677	-0.0483	4.60	0.0319
SEX		1	-0.4971	0.2675	-1.0214	0.0272	3.45	0.0631
PERIOD	1	0	0.0000	0.0000	0.0000	0.0000	.	.
PERIOD	2	1	0.4550	0.2715	-0.0771	0.9871	2.81	0.0937
PERIOD	3	1	-0.5386	0.4849	-1.4890	0.4119	1.23	0.2667
CD8GRP	1	1	-0.2150	0.6221	-1.4343	1.0044	0.12	0.7297
CD8GRP	2	1	-0.3646	0.7648	-1.8636	1.1345	0.23	0.6336
CD8GRP	3	0	0.0000	0.0000	0.0000	0.0000	.	.
CD8GRP	4	1	-0.3270	1.1595	-2.5996	1.9455	0.08	0.7779
CD8GRP	5	1	-0.8264	0.6057	-2.0136	0.3608	1.86	0.1725
Scale		0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
TRTSTATUS	1	4.48	0.0342
SEX	1	3.30	0.0695
PERIOD	2	5.54	0.0628
CD8GRP	4	5.35	0.2528

Модель, включающая CD8, учитываемая как серия индикаторных переменных Глава 33

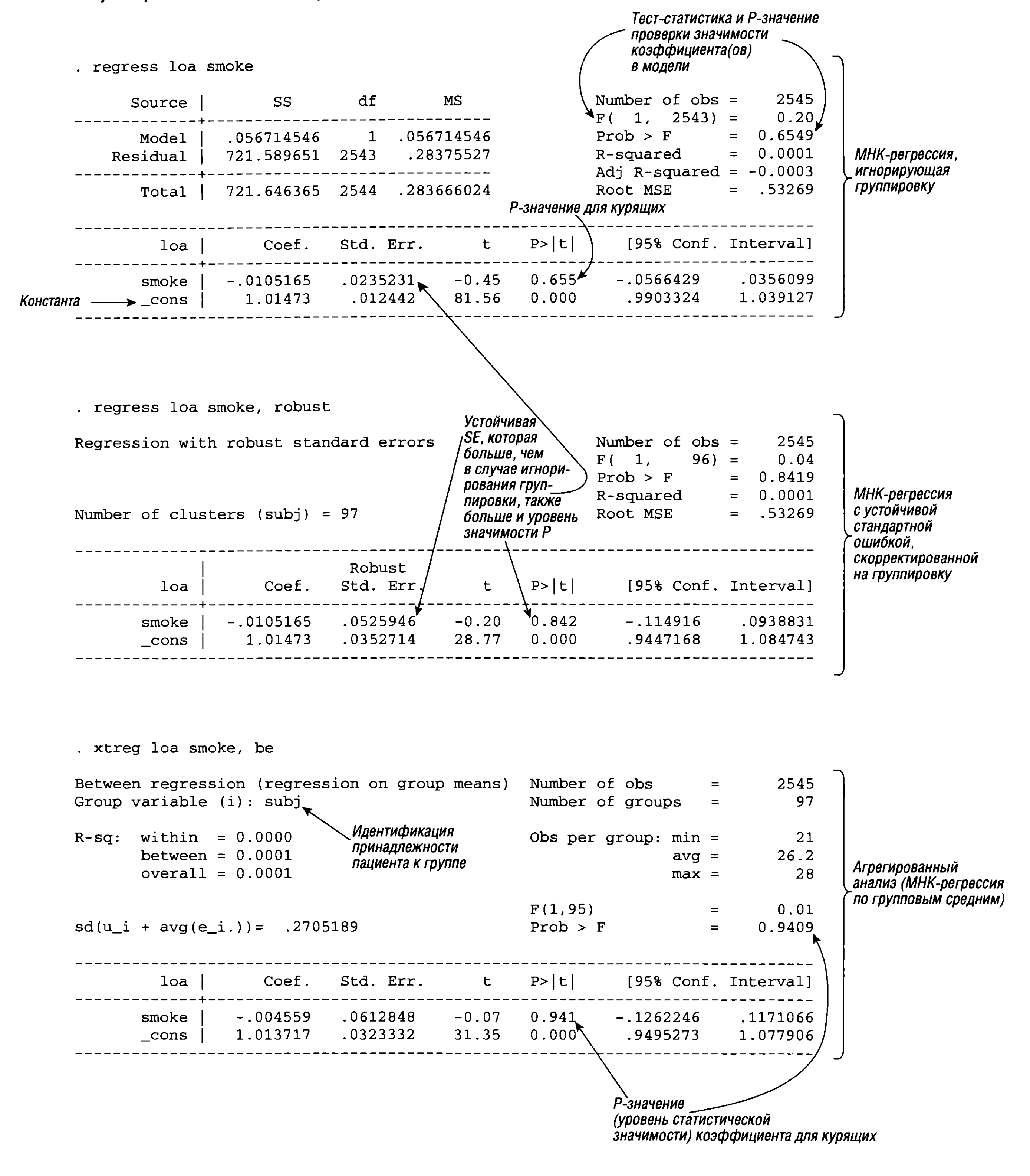
Параметры оценок для индикаторных переменных для исходного уровня CD8, учитывающих три градации (≥25, <1100) описанных категорий

Число дополнительных переменных в большей модели

Тест-статистика =392,5001–387,1458

Р-значение для критерия значимости исходного уровня CD8, учитываемой как качественная переменная

Analysis of periodontal data used in Chapter 42, generated by Stata



```
. regress loa smoke

      Source |       SS       df       MS              Number of obs =    2545
-------------+------------------------------           F(  1,  2543) =    0.20
       Model |  .056714546     1  .056714546           Prob > F      =  0.6549
    Residual |  721.589651  2543   .28375527           R-squared     =  0.0001
-------------+------------------------------           Adj R-squared = -0.0003
       Total |  721.646365  2544  .283666024           Root MSE      =  .53269

------------------------------------------------------------------------------
         loa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -.0105165   .0235231    -0.45   0.655    -.0566429    .0356099
       _cons |    1.01473    .012442    81.56   0.000     .9903324    1.039127
------------------------------------------------------------------------------


. regress loa smoke, robust

Regression with robust standard errors                 Number of obs =    2545
                                                       F(  1,    96) =    0.04
                                                       Prob > F      =  0.8419
                                                       R-squared     =  0.0001
Number of clusters (subj) = 97                         Root MSE      =  .53269

------------------------------------------------------------------------------
             |               Robust
         loa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -.0105165   .0525946    -0.20   0.842     -.114916    .0938831
       _cons |    1.01473   .0352714    28.77   0.000     .9447168    1.084743
------------------------------------------------------------------------------


. xtreg loa smoke, be

Between regression (regression on group means)  Number of obs      =      2545
Group variable (i): subj                        Number of groups   =        97

R-sq:  within  = 0.0000                         Obs per group: min =        21
       between = 0.0001                                        avg =      26.2
       overall = 0.0001                                        max =        28

                                                F(1,95)            =      0.01
sd(u_i + avg(e_i.))=  .2705189                  Prob > F           =    0.9409

------------------------------------------------------------------------------
         loa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |   -.004559   .0612848    -0.07   0.941    -.1262246    .1171066
       _cons |   1.013717   .0323332    31.35   0.000     .9495273    1.077906
------------------------------------------------------------------------------
```

```
. iis subj
. xtreg loa smoke, pa robust corr(exchangeable)

Iteration 1: tolerance = .00516018
Iteration 2: tolerance = 2.204e-07

GEE population-averaged model                   Number of obs      =      2545
Group variable:                       subj      Number of groups   =        97
Link:                             identity      Obs per group: min =        21
Family:                           Gaussian                     avg =      26.2
Correlation:                  exchangeable                     max =        28
                                                Wald chi2(1)       =      0.01
Scale parameter:                  .2835381      Prob > chi2        =    0.9198

                              (standard errors adjusted for clustering on subj)
------------------------------------------------------------------------------
             |             Semi-robust
         loa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -.0053018   .0526501    -0.10   0.920    -.1084941    .0978905
       _cons |   1.013841   .0347063    29.21   0.000     .9458185    1.081865
------------------------------------------------------------------------------
```

Корреляционная структура идентифицируется как взаимозаменяемая

Модель Хи-квадрат для тестирования значимости коэффициента в модели

P-значение для модели хи-квадрат

Тест-статистика Вальда

Обобщённые оценки уравнения (ООУ) с устойчивой стандартной ошибкой и взаимозаменяемой корреляционной структурой

```
. xtreg loa smoke, mle

Fitting constant-only model:
Iteration 0:   log likelihood = -1785.7026
Iteration 1:   log likelihood = -1785.7004

Fitting full model:
Iteration 0:   log likelihood = -1785.7027
Iteration 1:   log likelihood = -1785.6966
Iteration 2:   log likelihood = -1785.6966

Random-effects ML regression                    Number of obs      =      2545
Group variable (i): subj                        Number of groups   =        97

Random effects u_i ~ Gaussian                   Obs per group: min =        21
                                                               avg =      26.2
                                                               max =        28

                                                LR chi2(1)         =      0.01
                                               Prob > chi2        =    0.9302
------------------------------------------------------------------------------
         loa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -.0053168   .0607203    -0.09   0.930    -.1243265    .1136928
       _cons |   1.013844    .032046    31.64   0.000      .951035    1.076653
-------------+----------------------------------------------------------------
    /sigma_u |   .2519226   .0204583    12.31   0.000     .2118251    .2920201
    /sigma_e |   .4684954   .0066952    69.98   0.000     .4553731    .4816176
-------------+----------------------------------------------------------------
         rho |   .2242953   .0288039                      .1719879    .2846119
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=  443.21 Prob>=chibar2 = 0.000
```

Разность = –0,0038, так что –2×log(отношение правдоподобия)=2×0,0038=0,0076≈ 0,01

Конечная итерация, обеспечивающая стабильность оценок

СОП (статистика отношения правдоподобия) = –2×log (отношение правдоподобия)

Степени свободы

P-значение

log (правдоподобия) = –1785,6966

σ_c → /sigma_u

σ → /sigma_e

Модель случайных эффектов

Внутригрупповой коэффициент корреляции

$$= \frac{0{,}2519226^2}{0{,}2519226^2 + 0{,}4684954^2}$$

Приложение Г. Словарь терминов

2×2 таблица — таблица сопряжённости частот с двумя строками и двумя колонками.

−2×log(правдоподобие) — см. статистика отношения правдоподобия.

ANOVA — см. дисперсионный анализ.

ASCII, или текстовый формат — данные в компьютере даны как ряды текста.

C-статистика: описывает площадь под ROC-кривой и может использоваться, чтобы сравнивать диагностические тесты на одно и то же состояние.

CONSORT требования — формуляр, обеспечивающий критическую оценку и интерпретацию рандомизированных клинических испытаний (исследований), в форме контрольного списка и структурной схемы, перечисляющих, какую именно информацию должны сообщить авторы исследования.

***F*-критерий** — см. критерий отношения дисперсий.

***F*-распределение** — вытянутое вправо непрерывное распределение, характеризующееся степенями свободы числителя и знаменателя отношения, которое определяет его; полезно для сравнения двух дисперсий и более двух средних, применяется в дисперсионном анализе.

***P*-значение** — вероятность получения наших результатов или чего-либо большего, если нулевая гипотеза верна; уровень значимости.

R^2 — доля общей дисперсии зависимой переменной в регрессионном анализе, которая объясняется моделью; субъективное измерение качества согласия.

Автоматический выбор модели — метод выбора переменных для включения в модель, например, включение (пошаговое), исключение, пошаговое включение с исключением, все поднаборы переменных.

Альтернативная гипотеза — гипотеза относительно интересующего нас эффекта, которая противоречит нулевой гипотезе и верна, если нулевая гипотеза ложная.

Анализ выживаемости: исследует время, регистрируемое у пациента, чтобы достичь им интересующей нас конечной точки (например, смерть), когда некоторые данные цензурированы.

Анализ по назначению лечения — клиническое исследование, при котором все пациенты анализируются в тех группах, в которые они были назначены вначале.

Анализ пролеченных пациентов — клиническое исследование, в котором пациенты включаются в анализ, только если они закончили полный курс лечения, на который они были назначены случайным образом.

Априорная вероятность — индивидуальная уверенность, основанная на субъективной точке зрения и/или ретроспективных наблюдениях, что событие произойдёт.

Априорная (доопытная) вероятность — априорная вероятность, оценённая до появления результата диагностического теста, что пациент болен.

Апостериорная (послеопытная) вероятность — индивидуальное доверие, основанное на априорной вероятности и новой информации (например, результат критериальной проверки), в то, что событие произойдёт.

Апостериорные уточняющие сравнения: делаются для корректировки величин *P*, когда проведены множественные сравнения, например метод Бонферрони.

Асимметричное распределение — асимметричное распределение данных; оно имеет длинный хвост справа с несколькими высокими значениями (положительно скошенное) или длинный хвост слева с несколькими низкими значениями (отрицательно скошенное).

Арифметическое среднее — мера положения, полученная делением суммы значений переменной по наблюдениям на число слагаемых, часто называемая просто средним.

Байесовский метод вывода: применяет не только текущую информацию (т.е. от исследования), но и прежнее суждение субъекта (часто субъективное) о гипотезе, для оценки апостериорного доверия к гипотезе.

Бимодальное распределение — данные, чьё распределение имеет два «пика».

Бинарная переменная — качественная переменная с двумя категориями, также называется «дихотомическая переменная».

Биномиальное распределение — распределение дискретной вероятности бинарной случайной переменной; полезно для заключений о пропорциях.

Блок, группа — однородная группа экспериментальных единиц, которые имеют подобные характеристики, также называется «страта».

Бонферрони коррекция — коррекция значения уровня статистической значимости *P*, учитывающая число проверяемых гипотез, выполняемых при многократном испытании гипотез.

Бутстрэпинг — процесс моделирования, используемый обычно при получении оценок доверительного интервала для параметра. Использует для оценки параметра многократное извлечение случайных выборок, полученных из первоначальной выборки; после получения доверительного интервала рассматривается вариабельность распределения этих оценок.

Британский стандарт коэффициента воспроизводимости — максимальная разность, которая, по-видимому, существует между, максимум двумя, повторными измерениями.

Вариация остатков — вариация переменной, которая остаётся после того, как удалена вариабельность, относящаяся к интересующим нас факторам; вариация, не объяснимая моделью, и остаточный средний квадрат в таблице ANOVA. Также называется «ошибочная, или необъяснённая, вариация».

Взаимодействие: существует между двумя или более факторами, где разности зависимого признака между уровнями одного фактора различны для одного или более уровней другого фактора.

Взаимоисключающие категории (градации, уровни) — каждый пациент может принадлежать только к одной категории.

Взвешенная каппа — корректировка каппы Кохена, показателя согласия измерений, которое принимает во внимание степень, до которой расходятся две группы парных категориальных измерений.

Взвешенное среднее — модификация среднего арифметического, полученная путём учёта веса по каждому значению переменной в группе данных.

Весь набор моделей выбора — см. выбор модели.

Вероятность: измеряет степень возможности появления события. Она лежит между 0 и 1. См. также условная, априорная и апостериорная вероятность.

Воспроизводимость — степень, до которой совпадают повторные измерения того же наблюдателя в идентичных условиях.

Временной ряд — значения переменной, наблюдаемые у пациента или группы пациентов во многих последовательных точках во времени.

Вторая конечная точка — исходы в клиническом исследовании, которые не имеют главного значения.

Выбор модели — процедура формирования упрощённой схемы изучаемого явления. В регрессионном анализе выбор модели может проводиться с использованием разных алгоритмов отбора предикторов в уравнении регрессии. Наиболее популярны модели пошагового включения и исключения предикторов, а также их комбинации.

Выборка — подгруппа пациентов.

Выборочная квота — неслучайная выборка, в которой исследователь выбирает членов выборки для выполнения определённой «квоты».

Выборочное распределение средних — распределение выборочных средних, полученных после взятия повторных выборок фиксированного размера из популяции.

Выборочное распределение пропорций (долей) — распределение пропорций выборки, полученных после взятия повторных выборок фиксированного размера из популяции.

Выброс (аномальное значение) — наблюдение, которое отличается от основной части данных и несовместимо с остальными данными.

Гармонический анализ — временная серия, которая представлена суммой синусов и косинусов, предварительно очерченных для заданного периода и амплитуды.

Гауссово распределение — см. нормальное распределение.

Гетерогенность дисперсий — неравные генеральные дисперсии.

Гистограмма — диаграмма, которая показывает (относительное) распределение частоты непрерывной переменной путём применения присоединённых столбчатых диаграмм. Область столбчатой диаграммы пропорциональна (относительной) частоте в интервале, определённом границами этой диаграммы.

Гомоскедастичность — равные дисперсии; также описываются как гомогенность дисперсии.

График «стебель и листья» — смесь диаграммы и таблицы, применяемая для иллюстрации распределения данных. Он подобен гистограмме и эффективно отображает значения данных в возрастающем порядке.

График нормального распределения — диаграмма для визуальной оценки нормальности распределения данных; прямая линия на таком графике означает нормальность.

График «ящик с усами» — диаграмма, показывающая распределение переменной; она показывает средние, верхние и нижние квартили и часто максимальное и минимальное значения.

Группа отрицательного контроля (сравнения) — пациенты в рандомизированном контролируемом исследовании, которые не получают активного лечения.

Групповая рандомизация — группы пациентов, а не отдельные пациенты, случайным образом назначенные на лечение.

Групповое рандомизированное исследование — исследование, в котором группы пациентов рандомизированы по различным методам лечения так, что каждый пациент в пределах отдельной группы получает то же самое лечение.

Данные наблюдений — наблюдения над одной или более переменными.

Данные, цензурированные справа: относятся к пациентам, которые, как известно, не достигли интересующей нас конечной точки, когда они прослеживались в последний раз.

Дважды «слепой» (опыт, исследование) — см. «ослепление».

Двухмерный график (скаттер-плот) — двухмерный график зависимости одной переменной от другой, причём каждая пара значений отмечена точкой (наблюдением).

Двухвыборочный *t*-критерий — см. непарный *t*-критерий.

Двухвыборочный критерий Вилкоксона ранговых сумм — непараметрический критерий, сравнивающий распределения двух независимых групп наблюдений. Он эквивалентен U-критерию Манна–Уитни.

Двусторонний критерий — направление (знак) интересующего нас эффекта в альтернативной гипотезе не определено.

Децили — величины, которые делят упорядоченные наблюдения на 10 равных частей (по числу наблюдений).

ДИ — см. доверительный интервал.

Диагностический тест: используется для постановки диагноза особого условия.

Дисперсионный анализ (ANOVA) — общий термин для методов, которые сравнивают средние значения групп наблюдений путём расщепления общей дисперсии переменной на её компоненты, причём каждая относится к отдельному фактору.

Дисперсия — мера рассеяния, равная квадрату стандартного отклонения.

Дисперсия ошибок — см. дисперсия остатков.

Дискретное распределение вероятностей — случайная переменная, определяющая, как распределение принимает дискретные значения.

Дискретная переменная — числовая переменная, которая может принимать только целые числа.

Дискриминантный анализ — метод, подобный логистической регрессии, который можно применять для идентификации факторов, которые существенно связаны с бинарным ответом (группирующим признаком).

Дихотомическая переменная — см. бинарная переменная.

Доверительные границы — верхняя и нижняя величины доверительного интервала.

Доверительный интервал для параметра — диапазон значений, внутри которого, как мы (обычно) верим на 95%, лежит истинный параметр популяции. Строго говоря, после повторных отборов в этом интервале лежит 95% оценок этого параметра.

Доказательная медицина — применение текущих лучших результатов при принятии решения о лечении отдельных пациентов.

Зависимая переменная — переменная (обычно обозначенная как *Y*), которая предсказана независимой переменной в регрессионном анализе. Также называется откликом или зависимой выхода.

Заключение, вывод — процесс выведения заключения о популяции с применением выборочных данных.

Знаковый критерий — непараметрический критерий, который изучает, имеют ли тенденцию разницы быть положительными (или отрицательными); стремятся ли наблюдения быть больше (или меньше), чем средняя; больше (или меньше) одной пропорция наблюдений с определённой характеристикой, чем вторая.

«Золотой стандарт» (референтный, эталонный метод): даёт определённый диагноз конкретного состояния.

Изучение наблюдением — исследователь не делает ничего, чтобы влиять на исход.

Интервальная оценка — интервал значений, внутри которого, как мы полагаем, лежит параметр популяции.

Интердецильный размах — интервал между 10-й и 90-й процентилями; он содержит центральные 80% упорядоченных значений.

Интерполировать — оценить требуемое значение, которое лежит между двумя известными значениями.

Искажающие, вмешивающиеся (факторы) — положение в исследовании, когда одна или более независимых переменных соотносятся с исходом и друг с другом, так что трудно оценить независимый эффект каждой на переменную исхода.

Исследование «случай–контроль» — исследование, в котором идентифицируются группы пациентов с заболеванием (случаи) и без него (контроль) и сравнивается подверженность факторам риска в этих группах.

Искусственная переменная — группа бинарных переменных, которые созданы для облегчения сравнения трёх или более категорий номинальной переменной в регрессионном анализе.

Исторический, или непараллельный, контроль (пациентов) — использование пациентов, которые не отнесены к группе лечения в начале исследования, но получали лечение в прошлом, в качестве группы сравнения.

Каппа Кохена — мера согласия между двумя группами качественных измерений на одних субъектах. Если κ=1 — совершенное согласие, если κ=0 — не лучше, чем случайное согласие.

Качественная переменная — см. категориальная переменная.

Категориальная (качественная) переменная — каждый пациент принадлежит к одной из определённых категорий переменной.

Качество согласия — мера степени, до которой значения, получаемые из модели, согласуются с наблюдаемыми данными.

Квартили — значения, которые делят упорядоченные наблюдения на четыре равные части.

Клетка таблицы сопряжённости — пересечение отдельного ряда и отдельной колонки таблицы.

Клиническая гетерогенность: существует, когда исследования, включённые в метаанализ, имеют различия в популяции пациентов, определении переменных и т.п., что создаёт проблемы несовместимости.

Клиническое испытание — любая форма планируемого эксперимента на людях, который применяется для оценки влияния нового вида лечения на клинический исход.

Клиническая когорта — группа пациентов с одним и тем же клиническим состоянием, чьи исходы наблюдаются через какое-то время.

Ковариата — см. независимая переменная.

Ковариационный анализ — специальная форма анализа дисперсии, которая сравнивает значения зависимой переменной между группами пациентов, после корректировки эффекта одной или двух независимых переменных.

Контрольная (проверочная) выборка — вторая подвыборка, применяемая для удостоверения результатов от обучающей выборки.

Кохрановское сотрудничество — международная сеть клиницистов, методологов и потребителей, которые постоянно модернизируют систематические обзоры и делают их доступными для других.

Когортное исследование: прослеживается (обычно проспективно) группа пациентов, причём все без интересующего нас исхода (например, заболевание), чтобы изучить эффект будущих исходов на подверженность фактору риска.

Количественная переменная — см. числовая переменная.

Коллинеарность — пары независимых переменных в регрессионном анализе очень высоко коррелируют, т.е. с коэффициентами корреляции, очень близкими к ±1.

Конфаундинг — зависимость одной или более объясняющих переменных как от результирующей переменной, так и друг от друга, в результате чего трудно оценить независимый эффект каждой переменной на результирующую.

Копирование — случай, при котором у пациента есть больше одного измерения переменной.

Коррелограмма — двухмерный график коэффициента автокорреляции лага k.

Коэффициент вариации — стандартное отклонение, делённое на среднее (часто выражено в процентном отношении).

Коэффициент корреляции Пирсона — количественная мера между −1 и +1, отражает степень, до которой точки в двухмерном графике согласуются с прямой линией. См. также коэффициент корреляции Спирмэна.

Контроль — исследуемый пациент без заболевания в исследовании «случай–контроль» или не принимающий новое лечение в клиническом исследовании.

Контрольная группа — термин, применяемый в сравнительных исследованиях, например, в клинических испытаниях, для обозначения группы сравнения. См. также положительный и отрицательный контроль.

Коэффициенты регрессии — параметры (например, наклон и пересечение в парной регрессии), которые описывают уравнение регрессии.

Коэффициент логистической регрессии — коэффициент регрессии для конкретного предиктора в логистической регрессии.

Коэффициент ранговой корреляции — см. коэффициент корреляции Спирмэна.

Кривая выживаемости Каплана–Майера — кривая, в которой вероятность выживаемости наносится относительно времени от базовой величины. Используется, когда точные времена для достижения конечной точки известны.

Кривая операционной характеристики, она же характеристическая кривая теста — двухмерный график зависимости теста от величины (1−специфичность теста) для различных значений точки отсечки для непрерывной переменной в диагностическом тесте; применяется для отбора оптимальной величины отсечки или для сравнения тестов.

Критерий Бартлетта: применяется для сравнения дисперсий (генеральных).

Критерий знаковых рангов Вилкоксона — непараметрический критерий, сравнивающий парные наблюдения.

Критерий значимости — см. критерий проверки гипотез.

Критерий проверки гипотез — процесс использования выборки, чтобы оценить, как много аргументов имеется против нулевой гипотезы о популяции. Также называется «критерий значимости».

Критерий, свободный от распределения, — см. непараметрический критерий.

Критерий Колмогорова–Смирнова: определяет, нормально ли распределены данные.

Критерий Краскела–Уоллиса — непараметрическая альтернатива однофакторному ANOVA; применяется для

сравнения распределений более двух независимых групп наблюдений.

Критерий Левене: проверяет нулевую гипотезу, что две или более дисперсии равны.

Критерий Мак—Немара: сравнивает доли (пропорции) в двух соотносящихся группах, применяя статистику критерия Хи-квадрат Пирсона.

Критерий Манн—Уитни — см. критерий ранговых сумм.

Критерий отношения дисперсий — *F*-критерий Фишера—Снедекора, используется для проверки гипотез о равенстве дисперсий в популяции.

Критерий Хи-квадрат Пирсона: используется в частотных данных. Он проверяет нулевую гипотезу, что нет связи между факторами, которые определяют таблицу сопряжённости. Также применяется для тестирования разницы в долях (пропорциях).

Критерий Шапиро—Уилки: определяет, нормально ли распределены данные.

Круговой график — диаграмма, показывающая частотное распределение количественной или дискретной переменной. «Круг» делится на секции, одна для каждой «категории»: площадь каждой секции пропорциональна частоте в этой категории.

Линейное соотношение: означает прямолинейное соотношение между двумя переменными.

Ложноотрицательный — пациент, который имеет заболевание, но диагностируется как не имеющий его.

Ложноположительный — пациент, который не имеет заболевания, но диагностируется как имеющий его.

Логит- (логистическое) преобразование — преобразование, применённое к доле (пропорции) или вероятности, p, так что $\text{logit}(p)=\ln[p/(1-p)]$.

Логистическая регрессия — соотношение регрессии между бинарной переменной исхода и числом зависимых переменных.

Лог-нормальное распределение — вытянутое вправо распределение вероятности непрерывной случайной переменной, чей логарифм подчиняется нормальному распределению.

Лог-ранговый критерий — непараметрический подход к сравнению двух кривых выживаемости.

Лонгитюдальное (продольное) исследование: прослеживает пациентов в некоторый период времени.

Медиана — мера положения, срединное (расположенное посередине) значение упорядоченных наблюдений.

Межквартильный размах — интервал между 25-й и 75-й процентилями; он содержит центральные 50% упорядоченных значений.

Метаанализ (обзор) — количественный систематический обзор, который соединяет результаты исследований для создания и исследования оценки общего интересующего нас эффекта.

Метод наименьших квадратов (МНК) — метод оценки параметров в регрессионном анализе, основанный на минимизации суммы квадратов остатков.

Метод «складного ножа» — метод оценки параметров в модели: каждый из n пациентов последовательно удаляется из выборки, параметры оцениваются из остальных (n−1) пациентов, и, наконец, эти n оценки усредняются.

Многоуровневое моделирование — иерархическое расширение, применяемое для сложных структур в регрессионном анализе.

Множественная линейная регрессия — модель линейной регрессии, в которой одна зависимая переменная и две или более независимых переменных.

Мода — величина отдельной переменной, которая наиболее часто появляется в группе данных.

Модель: описывает в алгебраических терминах соотношение между двумя или более переменными.

Модель случайных эффектов: используется в метаанализе, когда есть данные о статистической гетерогенности.

Модель фиксированного эффекта: применяется в метаанализе, когда нет данных о статистической гетерогенности.

Модель Хи-квадрат (Хи-квадрат для предикторов) — статистика критерия, с распределением χ^2, которая проверяет нулевую гипотезу, что все отдельные коэффициенты регрессии в модели равны нулю.

Мощность (критерия) — вероятность отбрасывания нулевой гипотезы, когда она ложна.

Накопленная (нарастающим итогом) частота — число пациентов, которые имеют значения переменной меньше или равное определённому значению.

Независимая выборка — выборка, в которой каждое наблюдение представлено только однажды и не соотносится с наблюдениями в других выборках.

Независимая (объясняющая) переменная — переменная (обычно обозначаемая как x), которая применяется для прогноза зависимой переменной в регрессионном анализе. Также называется «независимая или предикторная переменная», или «ковариата».

Необъяснённая вариация — см. вариация остатков.

Непрерывная переменная — числовая переменная, не имеющая ограничения на значения, которые эта переменная может принимать.

Непрерывное распределение вероятностей — случайная переменная, определяющая распределение, непрерывна.

Несмещённый — свободный от смещения.

Номинальная переменная — категориальная переменная, чьи категории (уровни, градации) не имеют естественного упорядочения.

Номограмма Альтмана — диаграмма, которая устанавливает связь объёма выборки, мощности статистического критерия, уровень значимости и стандартизованную разность.

Номограмма Пагана — диаграмма, относящая предтестовую вероятность диагностического критерия к правдоподобию и посттестовой вероятности. Обычно используется для преобразования первой во вторую.

Нормальное (Гауссовское) распределение — непрерывное распределение вероятности, имеющее форму колокола и симметричное; его параметры — средняя и дисперсия.

Непарный (двухвыборочный) *t*-критерий: проверяет нулевую гипотезу, что две средних от независимых групп равны.

Непараметрический критерий — критерий проверки гипотез, который не делает предположений о распределении анализируемых данных. Иногда называется критерий, свободный от распределения, или ранговый метод.

Нулевая гипотеза (H_0) — утверждение, которое не предполагает обнаружения влияния вмешательства (лечения) в популяции.

Обзор — см. метаанализ.

Обоснованность — близость к истинному результату.

Обратный отбор — см. автоматический выбор модели.

Обучающая выборка — первая подвыборка, применяемая для генерирования модели (например, в анализе логистической регрессии или дискриминантном).

Результаты перепроверяются по второй (контрольной) выборке.

Одновыборочный t-критерий (Стьюдента): исследует, отличается ли средняя переменной от некоторой гипотетической величины.

Однократное маскирование («ослепление») — см. «ослепление».

Одномоментное (поперечное) исследование — исследование, которое проводится в отдельной точке во времени.

Односторонний критерий — альтернативная гипотеза, определяет направление интересующего нас влияния вмешательства.

Однофакторный дисперсионный анализ — отдельный вид ANOVA, применяемый для сравнения средних двух и более независимых групп наблюдений.

Ожидаемая частота — частота, которая ожидается исходя из предположения о правильности нулевой гипотезы.

Ординальная (ранговая) переменная — категориальная переменная, чьи категории расположены упорядоченным образом.

«Ослепление» — неосведомлённость в клиническом исследовании пациентов, клиницистов и консультантов о назначении лечения (двойное—слепое), или ситуация, когда пациент знает о получаемом лечении, а консультант — нет (одинарное—слепое).

Основная переменная исследования — переменная, которая относится к основной цели исследования.

Остатки — разность между наблюдаемыми и предсказанными значениями зависимой переменной в регрессионном анализе.

Остаточный эффект — остаточный эффект предыдущего вида лечения в перекрёстном исследовании.

Отбор (признаков) методом включения — см. модель выбора.

Отбор с последовательным удалением — см. выбор модели.

Относительная частота — частота, выраженная как процент или доля общей частоты.

Относительный шанс — соотношение двух шансов, интерпретируемое как относительный риск. Также называется «отношение шансов».

Относительный риск — соотношение двух рисков, обычно риск заболевания в группе пациентов, подверженных некоторому фактору, делённый на риск неподверженных пациентов.

Отношение дисперсий (F-критерий): применяется для сравнения дисперсий двух переменных путём сравнения их отношения с F-распределением.

Отношение правдоподобия — отношение двух правдоподобий; для диагностических тестов ОП — отношение шансов получения конкретного результата теста у пациентов, имеющих и не имеющих заболевание.

Отношение рисков — см. относительный риск.

Оценка (выборочная) — значение, полученное из выборки, которое используется для представления параметра популяции.

Отношение шансов — отношение двух шансов (например, шансы заболевания у пациентов, подверженные или не подверженные влиянию фактора). Часто принимается как оценка относительного риска в исследовании «случай—контроль».

Отрицательное предсказанное значение — доля пациентов с отрицательным результатом теста, которые не имеют заболевания.

Отступления от протокола — пациенты, которые входят в исследование, но не удовлетворяют критериям протокола.

Оценка риска — см. прогностический индекс.

Ошибка 2-го рода — неотбрасывание нулевой гипотезы, когда она ложна.

Ошибка, обусловленная выборкой, — разность, относимая к процедуре взятия образцов в выборку, для которых наблюдаемые значения отличаются от тех, что есть в популяции.

Ошибка 1-го рода — отбрасывание нулевой гипотезы, когда она верна.

Параллельное испытание — исследование, при котором каждый пациент получает только одно лечение.

Параметр — обобщающая величина (например, средняя, пропорция), которая характеризует распределение вероятности. Его значение относится к популяции.

Параметрический критерий — критерий проверки гипотез, который делает определённые вероятностные предположения о данных.

Парная линейная регрессия — прямолинейное соотношение между отдельной зависимой переменной и одной независимой переменной.

Парные наблюдения: относятся к ответам от подобранных пар пациентов или одного и того же пациента в двух различных условиях.

Парный t-критерий: проверяет нулевую гипотезу, что среднее значение группы разностей парных наблюдений равно нулю.

Переопределённые (избыточные) модели — модели, которые содержат слишком много объясняющих переменных.

Перекрёстные исследования — исследования, в которых каждый исследуемый пациент получает более одного вида лечения, одно за другим в случайном порядке.

Переменная — любая величина, которая изменяется.

Переменная отклика (зависимая переменная) — см. зависимая переменная.

Переопределённая модель — модель, в которой число переменных равно или больше числа пациентов.

Пересечение (свободный член) — значение зависимой переменной в уравнении регрессии, когда величина(ы) независимой(ых) переменной(ых) равна(ы) нулю.

Периодограмма — графическое изображение, применяемое в гармоническом анализе, части анализа временных серий.

План полной рандомизации — экспериментальные единицы, случайно предназначенные в группы лечения.

Плацебо — инертное «лечение», применяемое в клиническом исследовании, которое идентично по своим внешним проявлениям активному лечению. Оно удаляет эффект получения лечения из терапевтического сравнения.

Повторность (кратность) измерений — интересующая нас переменная измеряется на том же самом пациенте более чем в одной группе условий (т.е. по различным поводам).

Повторность измерений в дисперсионном анализе — специальная форма дисперсионного анализа, применяемая, когда числовая переменная измеряется у каждого члена группы пациентов по ряду различных поводов.

Подбор — процесс отбора пациентов, подобных между собой относительно переменных, влияющих на интересующий нас ответ.

Поправка на непрерывность — поправка, применяемая к статистике критерия для корректировки аппроксимации дискретного распределения непрерывным распределением.

Поправка Бонферрони — апостериорная поправка значения достигнутого уровня значимости P, принимающая

во внимание число проверок гипотез, проведённых в ситуациях тестирования множественных гипотез.

Поперечное исследование — исследование соотношений между заболеваниями (или другими признаками, относящимися к здоровью) и другими переменными в определённой популяции в определённый момент времени.

Полиномиальная регрессия — нелинейное (т.е. квадратичное, кубическое и т.п.) соотношение между зависимой переменной и независимой переменной.

Полностью рандомизированный план — план исследования, при котором экспериментальные единицы (наблюдения) назначаются случайно в группы лечения.

Популяция (генеральная совокупность) — вся группа пациентов, в изучении которых мы заинтересованы.

Положительный контроль — пациенты в РКИ, получающие какой-либо вид активного лечения, как основа сравнения для нового вида лечения.

Пошаговый отбор — см. модель отбора.

Правдоподобие: исходя из гипотезы, описывает достоверность наблюдаемого результата (например, от критерия), если гипотеза верна (например, болезнь есть).

Предсказанное значение — предсказанное значение переменной отклика в регрессионном анализе, соответствующее отдельным значениям независимой переменной.

Предварительный анализ — предварительно планируемые виды анализа в промежуточной фазе исследования.

Пределы согласия — в оценке повторяемости это размах величин, между которыми, как мы ожидаем, лежат 95% разницы между повторяемыми измерениями в популяции.

Предикторная переменная — см. независимая переменная.

Преобразование данных: получается взятием одного и того же математического преобразования (например, логарифма) для каждого наблюдения.

Преимущественная выборка — группа пациентов, относительно которой предполагается, что она является представительной для популяции, из которой она выбрана, тогда как на самом деле её выбор обусловлен не случайным методом, а мотивами удобства, преимущественно близкого расположения и т.п.

Пригодная выборка — группа пациентов, которая, как полагают, представляет популяцию, из которой она отобрана, но отобрана потому, что она «под рукой», а не отобрана случайно.

Процентная точка — процентиль распределения; он означает долю распределения, которая лежит справа (т.е. в правом хвосте), слева (в левом хвосте) или как слева, так и справа.

Период «размыва» — интервал между окончанием одного периода лечения и началом второго в перекрёстном исследовании. Он позволяет рассеяться остаточным эффектам первого лечения.

Прогностическая ценность положительного результата — доля пациентов с положительным результатом диагностического теста, имеющая заболевание.

Первичная конечная точка — исход, который наиболее точно отражает преимущество новой терапии в клиническом исследовании.

Прогностический индекс: оценивает правдоподобие того, что субъект болен. Также называется «индекс риска».

Пропорция (доля) — соотношение числа интересующих нас событий к общему числу событий.

Проспективное исследование — исследование, в котором пациенты прослеживаются вперед от какой-то точки во времени.

Протокол — полное письменное описание всех аспектов клинического исследования.

Процентили — величины, которые делят упорядоченные наблюдения на 100 равных частей.

Прямая линейной регрессии — прямая линия, нанесённая на двухмерный график, которая определяется алгебраическим выражением, соединяющим две переменные.

Публикационное смещение — тенденция журналов публиковать только статьи, которые содержат статистически значимые результаты.

Равномерное распределение: не имеет «пиков», поскольку каждое значение равновероятно.

Размах — разность между наибольшим и наименьшим значениями переменной.

Размах, отвечающий нормальному распределению, — см. референтный интервал.

Ранговый коэффициент корреляции Спирмэна — непараметрическая альтернатива коэффициенту корреляции Пирсона; он обеспечивает оценку связи между двумя переменными.

Ранговые методы — см. непараметрические критерии.

Рандомизация — назначение пациентов в группы лечения случайным (вероятностным) образом. Может быть стратифицированной (контролирующей эффект важных факторов) или блокированной (гарантируя почти равные по размеру группы лечения).

Рандомизированное контролируемое исследование (РКИ) — сравнительное клиническое исследование, в котором существует случайное назначение пациентов на вид лечения.

Ранее диагностированные пациенты — пациенты, которые диагностировались ранее.

Распределение вероятности — теоретическое распределение, которое описывается математической моделью. Оно показывает вероятности всех возможных значений случайной переменной.

Распределение Пуассона — распределение дискретной вероятности случайной переменной, представляющей число явлений, происходящих случайно и независимо с фиксированной средней частотой.

Распределение Стьюдента — см. *t*-распределение.

Распространённость (преваленс) — число (доля, пропорция) пациентов с болезнью в данной точке во времени (точечная распространённость) или внутри определённого интервала (периодическая распространённость).

Регрессионная модель пропорционального риска (Кокса): используется в анализе выживаемости для изучения одновременного эффекта зависимых переменных на выживаемость.

Регрессия к среднему — явление, посредством которого группа крайних результатов сопровождается результатами, которые в среднем менее экстремальны, т.е. высокие отцы имеют более низких (но всё-таки высоких) сыновей.

Ретроспективное исследование — пациенты выбраны, и исследуются факторы, которые произошли в их прошлом.

Референтный интервал — интервал значений (обычно центральных 95%) переменной, который обычно наблюдается у здоровых пациентов. Также называется «нормальный, или референтный, размах».

Риск — мгновенный риск достижения конечной точки в анализе выживаемости.

Риск заболевания — вероятность развития заболевания в установленный временной период.

Сезонная вариация — значения интересующей нас переменной систематически изменяются согласно времени года.

Сериальная корреляция — корреляция между наблюдениями во временных сериях и наблюдениями, отделёнными между собой фиксированным временным интервалом.

Симметричное распределение — данные сконцентрированы вокруг некоторой средней точки, а форма распределения влево от средней точки — зеркальное отражение формы справа от неё.

Систематическое размещение — пациенты в клиническом исследовании назначены на лечение систематизированным, а не случайным образом.

Систематический обзор — формализованный и строгий подход к комбинации результатов от всех уместных изучений подобных исследований по тому же самому состоянию здоровья.

Систематический метод выборки — выборка отбирается из популяции, применяя некоторый систематический метод, но не метод, основанный на случайности.

СКО (SD) — см. стандартное отклонение.

Случай — пациент с заболеванием при исследовании «случай–контроль».

Случайная выборка — каждая возможная выборка данного размера в популяции с равной вероятностью может быть выбрана.

Случайная серия — временная серия, в которой нет автокорреляции.

Случайная переменная — величина, которая может принимать любое из группы взаимоисключающих значений с данной вероятностью.

Случайная вариация — вариабельность, которую нельзя отнести к любому объяснённому источнику.

Смещение — систематическая разница между результатами, полученными из исследования, и истинным положением дел.

Смещение выбора — систематическое нарушение данных в результате способа, которым субъекты включаются в исследование.

Смещение вследствие ошибок памяти — систематическое искажение данных в результате характерной особенности, с помощью которой субъекты вспоминают прошлые события в отличие от событий недавних.

СОС (SEM) — см. стандартная ошибка среднего.

Стандартная ошибка среднего — мера точности выборочной средней; стандартное отклонение распределения средних отдельных выборок.

Среднее — общий термин для мер положения.

Среднее геометрическое — мера положения для данных, чьё распределение вытянуто вправо; антилогарифм средней арифметической для логарифмов исходных данных.

Среднее арифметическое — мера положения, полученная путём деления суммы значений переменной на число наблюдений. Часто называется «среднее».

Стандартное отклонение — стандартное (среднеквадратичное) отклонение; мера рассеяния, равная положительному квадратному корню из дисперсии.

Степени свободы (*df*) статистики обычно равны разности объёма наблюдений и единицы; вычисляются для конкретной статистики и указывают степень, в которой наблюдения являются «свободными», чтобы измениться.

Столбчатая диаграмма — диаграмма, которая показывает распределение качественной или дискретной переменной путём показа отдельной горизонтальной или вертикальной диаграммы для каждой «категории», причём её длина пропорциональна (относительной) частоте в этой «категории».

Структура выборки — список всех пациентов в популяции.

Специфичность (теста) — доля пациентов без заболевания, которые точно идентифицированы диагностическим тестом.

Стандартная ошибка пропорции (доли) — мера точности выборочной доли (пропорции); стандартное отклонение распределения выборки пропорций.

Стандартное нормальное распределение — специальное нормальное распределение со средним, равным нулю, и дисперсией, равной единице.

Стандартизованная разность — отношение, применяемое в номограмме Альтмана и формулах Лера, которое выражает клинически важную разность лечения в единицах стандартного отклонения.

Стандартизованное нормальное отклонение — случайная переменная, чьё распределение нормально с нулевой средней и единичной дисперсией.

Стационарный временной ряд — временная серия, для которой средняя и дисперсия постоянны всё время.

Статистика (как число) — выборочная оценка параметра популяции.

Статистическая гетерогенность присутствует в метаанализе, когда между отдельными оценками интересующего нас эффекта имеется значительная вариация.

Статистически значимый — результат проверки гипотезы статистическим критерием при определённом уровне (скажем, 1%), если есть существенные аргументы, чтобы отбросить нулевую гипотезу при этом уровне (т.е. когда $P < 0{,}01$).

Статистика (как наука): включает методы сбора, обобщения, анализа и вывода заключений из данных.

Страта — подгруппа пациентов: обычно пациенты внутри страты имеют подобные характеристики. Иногда называется «блок».

Субъективная вероятность — персональная степень уверенности, что событие произойдет.

Статистический критерий Вальда: часто применяется в логистической регрессии для проверки вклада отдельного коэффициента регрессии.

Статистика критерия — значение, выведённое из данных выборки, применяемое для проверки гипотез; это значение сравнивается с известным распределением вероятности (этого критерия) для получения величины p.

Сходимость (результатов) — степень, до которой те же результаты могут быть получены в различных условиях (т.е. двумя методами измерения или двумя наблюдателями).

***t*-распределение** — также называется t-распределением Стьюдента. Непрерывное распределение, чья форма подобна нормальному распределению и которое характеризуется своей степенью свободы. Оно особенно важно для заключений о средних.

Таблица 2×2 — таблица сопряжённости частот двух строк и двух колонок.

Таблица сопряжённости (контингенции) — двухвходовая (обычно) таблица, в которую вводятся частоты.

Таблицы продолжительности жизни в анализе выживаемости — способ определения вероятностей выживаемости, когда известно только, что время для достижения конечной точки находится в определённом временном интервале.

Теорема Байеса — апостериорная вероятность события/гипотезы пропорциональна произведению его априорной вероятности и правдоподобия.

Точечная оценка — отдельная величина, полученная из исследования, которая оценивает параметр популяции.

Точечная оценка распространённости (преваленса) — число пациентов с заболеванием (или процент уязвимых пациентов) в отдельной точке во времени.

Точный критерий Фишера — критерий, который оценивает точные вероятности (т.е. не полагается на аппроксимации по Хи-квадрат распределению Пирсона) в таблице сопряжённости (обычно в таблице 2×2), применяется, когда ожидаемые частоты невелики.

Точки влияния — значение данных, которое имеет эффект существенного изменения оценок коэффициентов регрессии, когда оно включено в анализ.

Точность: относится к способу, которым наблюдаемое значение количественно соответствует истинной величине.

Точность — мера ошибки выборки. Относится к тому, насколько хорошо совпадают друг с другом повторные наблюдения.

Точечный график — диаграмма, в которой каждое наблюдение за переменной представлено одной точкой на горизонтальной (или вертикальной) линии.

Тренд — значения переменной показывают тенденцию со временем прогрессивно увеличиваться или уменьшаться.

Угловой коэффициент (регрессии) — градиент линии регрессии, показывающий среднее изменение зависимой переменной на единицу изменения у независимой переменной.

Унимодальное распределение: имеет один «пик» (моду).

Уровень (градация) — конкретная категория качественной переменной или фактора.

Уровень значимости — вероятность, выбранная в начале исследования, которая приведёт нас к отбрасыванию нулевой гипотезы, если *P*-значение лежит ниже неё. Она часто выбирается как 0,05.

Уровень значимости апостериорного критерия — апостериорная вероятность, определённая из полученной ранее информации и результата диагностического теста, что пациент имеет заболевание.

Условная вероятность — вероятность события, при условии, что другое событие уже произошло.

Условная логистическая регрессия — вид логистической регрессии, когда пациенты в исследовании согласованы по конкретному условию.

Устойчивость (критерия) — критерий устойчив к нарушениям своих допущений, если на его величину *P* и мощность эти нарушения ощутимо не влияют.

Фактор риска — решающий фактор, который влияет на частоту особого исхода, например, заболевания.

Факторный эксперимент: разрешает одновременный анализ ряда интересующих нас факторов.

Форест-график — диаграмма, применяемая в метаанализе и показывающая оценённый эффект в каждом исследовании и их средние (с доверительными интервалами).

Формула Лера: может использоваться для расчёта оптимальных размеров выборки, требуемой для некоторых критериев проверки гипотез, когда мощность определена как 80% или 90%, а уровень значимости как 0,05.

Функция плотности вероятности — уравнение, которое определяет связь распределения вероятности и значения случайной величины.

Хи-квадрат критерий: используется на частотных данных. Он проверяет нулевую гипотезу о том, что нет никакой ассоциации между факторами, которые определяют таблицу сопряжённости. Также используется для проверки различий между пропорциями (долями).

Хи-квадрат (χ^2) **распределение Пирсона** — вытянутое вправо непрерывное распределение, характеризующееся степенями его свободы; полезно для анализа качественных данных.

Цензурированные данные: встречаются в анализе выживаемости, поскольку имеется неполная информация об исходе. См. цензурирование данных справа и слева.

Цензурирование данных слева: относится к пациентам, у которых наблюдение не начиналось до тех пор, пока не были получены базовые данные.

Циклическая изменчивость — значения показывают картину, которая повторяется после фиксированного периода.

Частная сумма (итог) в таблице сопряжённости — сумма частот в данном ряду (или колонке) таблицы.

Частота — число раз, когда происходит событие.

Частота новых случаев — число пациентов, которые заболевают в особый временной период, обычно выраженное как пропорция лиц, уязвимых в середине этого периода.

Частотная вероятность — доля, показывающая когда должно произойти событие, если бы мы повторяли эксперимент большее число раз.

Частотное распределение: показывает частоту появления каждого возможного наблюдения, класса наблюдений или категории.

Частный коэффициент регрессии — параметры, отличные от свободного члена (пересечения), которые описывают уравнение множественной линейной регрессии.

Число больных, которых необходимо лечить определённым методом в течение определённого времени, чтобы достичь определённого эффекта или предотвратить неблагоприятный исход (ЧБНЛ) — число пациентов, которым нужно экспериментальное, а не контрольное лечение, чтобы предупредить «плохой» исход одного из них.

Число вновь обнаруженных больных — пациенты, которым диагноз уже поставлен.

Числовая (количественная) переменная — переменная, которая принимает дискретные или непрерывные величины.

Чувствительность — доля пациентов с заболеванием, которые точно диагностированы тестом.

Шансы — соотношение вероятностей двух явлений, дополняющих друг друга (до 1); обычно вероятность иметь заболевание, делённое на вероятность не иметь его.

Экспериментальная единица — наименьшая группа пациентов, которая может рассматриваться для целей анализа.

Экспериментальное исследование — исследование, в которое учёный вмешивается каким-то образом для того, чтобы влиять на исход.

Эффект вмешательства — величина переменной отклика, которая отражает интересующее нас сравнение, например, разность в средних.

Эмпирическое распределение — наблюдаемое (фактическое) распределение переменной.

Эпидемиологическое исследование — наблюдательное исследование, которое оценивает соотношение между факторами риска и заболеванием.

Эффект здорового участника — при выборе пациентов без заболевания для участия в исследовании интересующий нас ответ (обычно смертность) в начале исследования ниже, чем можно было ожидать в общей популяции.

Эффект лечения — интересующий нас эффект (например, разность между средними относительного риска), который даёт сравнение лечения.

Эффект наложения — остаточный эффект прежнего измерения в перекрёстном исследовании.

Приложение к русскому изданию

БИБЛИОГРАФИЯ ОТ НАУЧНОГО РЕДАКТОРА

1. Автандилов Г.Г. Медицинская морфометрия. Руководство. — М.: Медицина, 1990. — 384 с.

2. Адлер Ю.П., Маркова Е.В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1976. — 279 с.

3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное издание. — М.: Финансы и статистика, 1983. — 471 с.

4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справочное издание. — М.: Финансы и статистика, 1985. — 487 с.

5. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. Справочное издание. — М.: Финансы и статистика, 1989. — 607 с.

6. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. — М.: Статистика, 1974.— 240 с.

7. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. — М.: Юнити, 1998. — 1022 с.

8. Александров В.В., Шнейдеров В.С. Обработка медико-биологических данных на ЭВМ. — Л.: Медицина, 1984. — 157 с.

9. Анализ данных на ЭВМ. (На примере системы СИТО) / Александров В.В., Алексеев А.И., Горский Н.Д. — М.: Финансы и статистика, 1990. — 192 с.

10. Андерсон Т. Статистический анализ временных рядов. — М.: Мир, 1976. — 755 с.

11. Андерсон Т. Введение в многомерный статистический анализ. — М., Физматгиз, 1963. — 500 с.

12. Аренс Х., Лёйтер Ю. Многомерный дисперсионный анализ. — М.: Финансы и статистика, 1985. — 230 с.

13. Аптон Г. Анализ таблиц сопряжённости. —. М.: Мир, 1982. — 126 с.

14. Архипова Г.П., Лаврова И.Г., Трошина И.М. Некоторые современные методы статистического анализа в медицине. — М.: РАМН, 1971. — 76 с.

15. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. Пер. с англ. — М.: Мир, 1982. — 488 с.

16. Банержи А. Медицинская статистика понятным языком: вводный курс. Пер. с англ. / Под ред. В.П. Леонова. — М.: Практическая медицина, 2007. — 207 с.

17. Бейли Н. Математика в биологии и медицине. — М.: Мир, 1970.— 270 с.

18. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов / Пер. с англ. — М.: Мир, 1971. — 408 с.

19. Берк К., Кэйри П. Анализ данных с помощью Microsoft Excel. Пер. с англ. — М.: Издательский дом «Вильямс», 2005. — 560 с.

20. Божко В.П., Хорошилова А.В. Информационные технологии в статистике. Учебник. — М.: Финстатинформ, 2002. — 144 с.

21. Бокс Дж., Дженкинс Г. Анализ временных рядов: Прогноз и управление. — М.: Мир, 1974.

22. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука, 1983. — 416 с.

23. Бондарь А.Г., Статюха Г.А., Потяженко И.А. Планирование эксперимента при оптимизации процессов химической технологии (Алгоритмы и примеры). — Киев: Вища школа, 1980. — 264 с.

24. Борель Э. Вероятность и достоверность. — М.: ГИФМЛ, 1961. — 120 с.

25. Боровиков В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. — 2-е издание. — СПб.: Питер, 2003 г. — 688 с.

26. Боровиков В. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере. — М.: Финансы и статистика, 2000 г. — 384 с.

27. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. — М.: Наука, 1983.— 464 с.

28. Браунли К.А. Статистическая теория и методология в науке и технике. Пер. с англ. — М.: Наука, 1977, — 407 с.

29. Бриллинджер Д. Временные ряды. Обработка данных и теория. — М.: Мир, 1980. — 536 с.

30. Броуди М. О статистическом рассуждении. — М.: Статистика, 1968. — 70 с.

31. Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. — СПб.: ДиаСофтЮП, 2002. — 608 с.

32. Вайнберг Дж., Шумекер Дж. Статистика. — М.: Статистика, 1979. — 390 с.

33. Вараксин А.Н. Статистические модели регрессионного типа в экологии и медицине. — Екатеринбург: Изд-во Гощинский, 2006. — 25 с.

34. Вейр Б. Анализ генетических данных. Пер с англ. — М.: Мир, 1995. — 400 с.

35. Вентцель Е.С. Теория вероятностей. — М.: Наука, 1969. — 576 с.

36. Вероятность и математическая статистика: Энциклопедия. — Научное издательство «Большая российская энциклопедия». 1999. — 910 с.

37. Владимирский Б.М. Математические методы в биологии. — Ростов: Изд-во Рост. ун-та, 1983. — 304 с.

38. Власов В.В. Введение в доказательную медицину. — М.: Медиа Сфера, 2001. — 392 с.

39. Власов В.В. Эпидемиология: Учеб. пос. для вузов. — М.: ГЭОТАР-МЕД, 2004. — 464 с. URL: http://www.biometrica.tomsk.ru/review3.htm

40. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. — М.: ФОРУМ: ИНФРА-М, 2004. — 464 с.

41. Вучков И. и др. Прикладной регрессионный анализ. — М.: Финансы и статистика, 1987. — 239 с.

42. Гарганеева Н.П. Леонов В.П. Логистическая регрессия в анализе связи артериальной гипертонии и психических расстройств // Сибирский медицинский журнал. — 2001. — № 3–4. — С. 42–48.— URL: http://www.biometrica.tomsk.ru/psycho3s.htm

43. Гарганеева Н.П. Леонов В.П. Возможности биометрического анализа взаимосвязи соматических показателей и систематики психических расстройств // Сибирский медицинский журнал. — 2001. — № 2. — С. 25–32. — URL: http://www.biometrica.tomsk.ru/psycho1.htm

44. Герасимович А.И. Математическая статистика. — Минск: Выш. школа, 1983. — 279 с.

45. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. Пер. с англ. — М.: Прогресс, 1976. — 495 с.

46. Гланц С. Медико-биологическая статистика. — М.: Практика, 1998. — 459 с.

47. Глотов Н.В., Животовский Л.А., Хованов Н.В., Хромов-Борисов Н.Н. Биометрия. — Л., 1982.

48. Гринхальх Т. Основы доказательной медицины. Пер. с англ. — М.: ГЭОТАР-МЕД, 2004. — 240 с. — URL: http://www.biometrica.tomsk.ru/review2.htm

49. Гублер Е.В. Информатика в патологии, клинической медицине и педиатрии. — Л.: Медицина, 1990. — 176 с.

50. Гублер Е.В., Генкин А.А. Применение непараметрических критериев статистики в медико-биологических исследованиях. — Л.: Медицина, 1973. — 141 с.

51. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. — Л.: Медицина, 1978. — 146 с.

52. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения. — М.: Мир, 1971.

53. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы планирования эксперимента. — М.: Мир, 1981. — 520 с.
54. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд. Пер. с англ. — М.: Издательский дом «Вильямс», 2007. — 912 с.
55. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. — М.: Финансы и статистика, 2000. — 352 с.
56. Дубровский С. А. Прикладной многомерный статистический анализ.— М.: Финансы и статистика, 1982.—216 с.
57. Дэвид Г. Метод парных сравнений. — М.: Статистика, 1978. — 144 с.
58. Дэниел К. Применение статистики в промышленном эксперименте. Пер. с англ. — М.: Мир, 1979. — 299 с.
59. Дюран Н., Оделл П. Кластерный анализ. — М.: Статистика, 1977. — 128 с.
60. Елисеева И.И., Рукавишникова В.О. Группировка, корреляция, распознавание образов (Статистические методы классификации и измерения связи). — М.: Статистика, 1977. — 144 с.
61. Елисеева И.И., Юзбашев М.М. Общая теория статистики. — М.: Финансы и статистика, 1998. — 368 с.
62. Ермолаев О.Ю. Математическая статистика для психологов. — М.: Изд-во Флинта, 2003. — 336 с.
63. Загоруйко Н. Г. Методы распознавания и их применение. — М.: Советское радио, 1972. — 288 с.
64. Закс Л. Статистическое оценивание. — М.: Статистика, 1976. — 598 с.
65. Зайцев Н.Г. Методика биометрических расчётов. — М.: Наука, 1973. — 256 с.
66. Зайцев Н.Г. Математическая статистика в экспериментальной ботанике. — М.: Наука, 1984. — 424 с.
67. Зорин Н.А. О неправильном употреблении термина достоверность в российских научных психиатрических и общемедицинских статьях. — URL: http://www.biometrica.tomsk.ru/let1.htm
68. Иберла К. Факторный анализ. — М.: Статистика, 1980. — 398 с.
69. Иванов Ю.И., Погорелюк О.Н. Обработка результатов медико-биологических исследований на микрокалькуляторах. — М.: Медицина, 1990. — 222 с.
70. Иванова И.М. Случайные числа и их применения. — М.: Финансы и статистика, 1984. — 111 с.
71. Кант В.И. Математические методы и моделирование в здравоохранении. — М.: Медицина, 1987. — 224 с.
72. Кендалл М., Стьюарт А. Теория распределений. — М.: Наука, 1966. — 588 с.
73. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973. — 899 с.
74. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. — М.: Наука, 1976. — 736 с.
75. Кендэл М. Ранговые корреляции. — М.: Статистика, 1975. — 214 с.
76. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. — М.: ФИЗМАТЛИТ, 2006. — 816 с.
77. Колемаев В.А. и др. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1991. — 400 с.
78. Коли Г. Анализ популяций позвоночных. Пер. с англ. — М.: Мир, 1979. — 362 с.
79. Компьютерная биометрика / Под ред. В.Н. Носова. — М.: Изд-во МГУ, 1990. — 232 с.
80. Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. — Минск: Изд-во БГУ, 1982. — 302 с.
81. Лакин Г.Ф. Биометрия. — М.: Высшая школа, 1990. — 352 с.
82. Леонов В.П. Обработка экспериментальных данных на программируемых микрокалькуляторах / Под ред. Б.А. Гладких. — Томск: Изд-во ТГУ, 1990. — 376 с.
83. Леонов В.П. Применение статистики в статьях и диссертациях по медицине и биологии. Часть III. Проблемы взаимодействия автор — редакция — читатель// Международный журнал медицинской практики. 1999. — Вып. 12. — С. 7–13. — URL: http://www.biometrica.tomsk.ru/naukoved/problem1.htm
84. Леонов В.П. Когда нельзя, но очень хочется, или Ещё раз о критерии Стьюдента. — URL: http://www.biometrica.tomsk.ru/student_1.htm
85. Леонов В.П. Наукометрика статистической парадигмы экспериментальной биомедицины (по материалам публикаций) // Вестник Томского государственного университета. Серия Математика. Кибернетика. Информатика. — № 275. — АПРЕЛЬ 2002. — С. 17–24. — URL: http://www.biometrica.tomsk.ru/paradigma.htm
86. Леонов В.П. Три «Почему...» и пять принципов описания статистики в биомедицинских публикациях. — URL: http://www.biometrica.tomsk.ru/principals.htm
87. Леонов В.П. Критерий Колмогорова–Смирнова: особенности применения. — URL: http://www.biometrica.tomsk.ru/k_s.htm
88. Леонов В.П. Кунсткамера. — URL: http://www.biometrica.tomsk.ru/kk.htm
89. Леонов В.П. Анализ таблиц сопряжённости с вычислением критерия Хи-квадрат и статистик связи. — URL: http://www.biometrica.tomsk.ru/frea.htm
90. Леонов В.П. Искушение Знанием, или Сколько стоит биометрика? — URL: http://www.biometrica.tomsk.ru/price_biom.htm
91. Леонов В.П. Обучение медиков статистике: попытка системного подхода к проблеме. — 2006. — Вып. 2. — С. 17–22. — URL: http://www.biometrica.tomsk.ru/leonov_2006.htm
92. Леонов В.П. Ошибки статистического анализа биомедицинских данных // Международный журнал медицинской практики. — 2007. — Вып. 2. — С. 19–35. — URL:http://www.biometrica.tomsk.ru/error.htm
93. Лиепа И.Я. Математические методы в биологических исследованиях. Факторный и компонентный анализы. — Рига, 1980. — 104 с.
94. Лисенков А.Н. Математические методы планирования многофакторных медико-биологических экспериментов. — М.: Медицина, 1979. — 344 с.
95. Лоули Д., Максвелл А. Факторный анализ как статистический метод. — М.: Мир, 1967. — 144 с.
96. Львовский Е.Н. Статистические методы построения эмпирических формул: Учеб. пособие. — М.: Высш. школа, 1982. — 224 с.
97. Любищев А.А. Дисперсионный анализ в биологии. — М.: Изд-во МГУ, 1986. — 200 с.
98. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учеб. пособие. — М.: Финансы и статистика, 2002. — 368 с.: ил.
99. Максимов В.Н. Многофакторный эксперимент в биологии. — М.: МГУ, 1980. — 279 с.
100. Максимов Г.К., Синицын А.Н. Статистическое моделирование многомерных систем в медицине. — М.: Медицина, 1983. — 144 с.
101. Малиновский Л.Г. Классификация объектов средствами дискриминантного анализа. — М.: Наука, 1979. — 260 с.
102. Мандель И.Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 176 с.
103. Методы современной биометрии: сб. статей / Отв. ред. Н.А. Плохинский. — М.: МГУ, 1978. — 207 с.
104. Мердок Дж. Контрольные карты. Пер. с англ. — М.: Финансы и статистика, 1986. — 132 с.
105. Меркурьева Е.К. Биометрия в селекции и генетике сельскохозяйственных животных. — М.: Колос, 1970. — 423 с.
106. Миркин Б.Г. Анализ качественных признаков. — М.: Статистика, 1976.— 168 с.
107. Миркин Б.Г. Анализ качественных признаков и структур. — М.: Статистика, 1980. — 320 с.
108. Митропольский А.К. Техника статистических вычислений. — М.: Наука, 1971. — 576 с.
109. Михок Г., Урсяну В. Выборочный метод и статистическое оценивание. — М.: Финансы и статистика, 1982. — 245 с.
110. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. — М.: Финансы и статистика, 1982. — 272 с.

111. Налимов В.В. Теория эксперимента. — М.: Наука, 1971. — 208 с.
112. Налимов В.В., Голикова Т.И. Логические основания планирования эксперимента. 2-е изд., перераб. и доп. — М.: Металлургия, 1980. — 152 с.
113. Налимов В.В., Чернова Н.А. Статистические методы планирования экстремальных экспериментов. — М.: Наука, 1965. — 340 с.
114. Неуймин Я.Г. Модели в науке и технике. История, теория, практика. — Л.: Наука, 1984. — 190 с.
115. Обучение медицинской статистике. Двадцать конспектов лекций и семинаров. Пер. с англ. Женева: ВОЗ, 1989. — 216 с.
116. Орлов А.И. Эконометрика. Учебник. — М.: Экзамен, 2002. — 576 с.
117. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. — М.: Мир, 1982. — 324 с.
118. Палий И.А. Прикладная статистика. — М.: Высшая школа, 2004. — 175 с.
119. Петри А., Сэбин К. Наглядная статистика в медицине. Пер. с англ. В.П. Леонова. — М.: ГЭОТАР-МЕД, 2003. — 144 с. — URL: http://www.biometrica.tomsk.ru/petrie.htm
120. Петрович М.Л. Регрессионный анализ и его математическое обеспечение на ЕС ЭВМ. — М.: Финансы и статистика, 1982. — 199 с.
121. Планирование эксперимента в биологии и сельском хозяйстве / Под ред. В.Н. Максимова. — М.: МГУ, 1991. — 220 с.
122. Плохинский Н.А. Математические методы в биологии. — М.: МГУ, 1978. — 226 с.
123. Плохинский Н.А. Алгоритмы биометрии. — М.: МГУ, 1980. — 150 с.
124. Поллард Дж. Справочник по вычислительным методам статистики. — М.: Финансы и статистика, 1982. — 344 с.
125. Применение математических методов в биологии. — Л.: ЛГУ, 1963. — 240 с.
126. Пустыльник Е.И. Статистические методы анализа и обработка наблюдений. — М.: Наука, 1968. — 288 с.
127. Пфанцагль И. Теория измерений. — М.: Мир, 1976. — 165 с.
128. Рао С. Линейные статистические методы и их применение. Пер. с англ. — М.: Наука, 1986. — 547 с.
129. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. — М.: Медиа Сфера, 2002. — 312 с. — URL: http://www.biometrica.tomsk.ru/review1.htm
130. Рекомендации по подготовке научных медицинских публикаций. Сборник статей и документов. — М.: Медиа Сфера, 2006. — 464 с.
131. Рокицкий П.Ф. Биологическая статистика. — Минск: Вышэйшая школа, 1973. — 320 с.
132. Ростова Н.С. Корреляции: структура и изменчивость. — СПб.: Изд-во С-Петерб. ун-та, 2002. — 308 с.
133. Рузинов Л.П. Статистические методы оптимизации химических процессов. — М.: Химия, 1972. — 200 с.
134. Рузинов Л.П., Слободчикова Р.И. Планирование эксперимента в химии и химической технологии. — М.: Химия, 1980. — 280 с.
135. Рунион Р. Справочник по непараметрической статистике. Современный подход. — М.: Финансы и статистика, 1982. — 198 с.
136. RusDASL — российская библиотека данных для изучающих биостатистику. — URL: http://www.biometrica.tomsk.ru/rus_dasl.htm
137. Саката Сиро. Практическое руководство по управлению качеством. Пер. с япон. — М.: Машиностроение, 1980. — 215 с..
138. Себер Дж. Линейный регрессионный анализ. Пер. с англ. — М.: Мир. — 456 с.
139. Сепетлиев Д. Статистические методы в научных медицинских исследованиях. — М.: Медицина, 1968. — 419 с.
140. Сергиенко В.И., Бондарева И.Б. Математическая статистика в клинических исследованиях. — М.: ГЭОТАР-МЕД, 2001. — 256 с.
141. Сильвестров Д.С. Программное обеспечение прикладной статистики. — М.: Финансы и статистика, 1988. — 240 с.
142. Снедекор Дж. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. — М.: Сельхозгиз, 1961. — 503 с.
143. Славин М.Б. Методы системного анализа в медицинских исследованиях. — М.: Медицина, 1989. — 304 с.
144. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1965. — 511 с.
145. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. — М.: Статистика, 1980. — 208 с.
146. Соколов Д.К. Математическое моделирование в медицине. — М.: Медицина, 1974. — 185 с.
147. Сошникова Л.А. и др. Многомерный статистический анализ в экономике: Учебное пособие для вузов. — М.: ЮНИТИ-ДАНА, 1999. — 598 с.
148. Справочник по прикладной статистике: В 2 т. Т. 1. Пер. с англ. / Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. — М.: Финансы и статистики, 1989. — 510 с. — Т. 2. — 1990. — 52 с.
149. Статистические методы для ЭВМ / Под ред. К. Энслейна, Э. Рэлстона, Г.С. Уилфа. — М.: Наука. 1986. — 464 с.
150. Суппес П., Зинес Дж. Основы теории измерений // Психологические измерения. — М.: Мир, 1967. — С. 9–110.
151. Терентьев П.В., Ростова Н.С. Практикум по биометрии. — Л.: ЛГУ, 1977. — 152 с.
152. Терехина А. Ю. Анализ данных методами многомерного шкалирования. — М.: Наука, 1986. — 168 с.
153. Трофимов В.П. Логическая структура статистических моделей. — М.: Финансы и статистика, 1985. — 191 с.
154. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. — М.: Финансы и статистика, 1995. — 384 с.
155. Уилкс С. Математическая статистика. — М.: Наука, 1967. — 632 с.
156. Урбах В.Ю. Биометрические методы. Статистическая обработка опытных данных в биологии, сельском хозяйстве и медицине. — М.: Наука, 1964. — 416 с.
Ким Дж.О., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ. Пер. с англ. / Под ред. И.С. Енюкова. — М.: Финансы и статистика, 1989. — 215 с.
157. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа. — М.: Финансы и статистика, 1983. — 302 с.
158. Финни Д. Введение в теорию планирования экспериментов. Пер. с англ. — М.: Наука, 1970. — 287 с.
159. Флейс Дж. Статистические методы для изучения таблиц долей и пропорций.: Пер. с англ. — М.: Финансы и статистика, 1981. — 319 с.
160. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. — М.: Мир, 1989. — 512 с.
161. Хан Д. Планирование и контроль: концепция контроллинга. Пер. с нем. — М.: Финансы и статистика, 1997. — 800 с.
162. Хан Г., Шапиро С. Статистические модели в инженерных задачах. — М., Мир, 1969. — 395 с.
163. Харман Г. Современный факторный анализ. Пер. с англ. В.Я. Лумельского / Научное редактирование и вступительная статья Э.М. Бравермана. — М.: Статистика, 1972. — 486 с.
164. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. Пер. с англ. А.К. Звонкина. — М.: Статистика, 1980. — 95 с.
165. Хеннан Э. Многомерные временные ряды. — М.: Мир, 1974.
166. Химмельблау Д. Анализ процессов статистическими методами. Пер. с англ. — М.: Мир, 1973. — 957 с.
167. Холлендер М., Вулф Д. Непараметрические методы статистики. Пер. с англ. Д.С. Шмерлинга. — М.: Финансы и статистика, 1983. — 518 с.
168. Хургин Я.И. Да, нет или может быть... — 2-е изд. — М.: Наука, 1983. — 208 с.

11 Заказ 2051

169. Хургин Я.И. Как объять необъятное. — М.: Знание, 1979. — 192 с.
170. Хьюбер П. Робастность в статистике. — М.: Мир, 1984. — 304 с.
171. Хикс Ч. Основные принципы планирования экспериментов. — М.: Мир, 1967. — 406 с.
172. Четыркин Е.М., Калихман И.Л. Вероятность и статистика. — М.: Финансы и статистика, 1982. — 319 с.
173. Четыркин Е.М. Статистические методы прогнозирования. — М.: Статистика, 1977.
174. Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. Пер. с нем. — М.: Мир, 1970. — 368 с.
175. Шмидт-Ниельсен К. Размеры животных: почему они так важны? Пер. с англ. — М.: Мир, 1987. — 259 с.
176. Эльясберг П.Е. Измерительная информация. Сколько её нужно, как её обрабатывать? — М.: Наука, 1983. — 208 с.
177. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. — М.: Финансы и статистика, 1988. — 263 с.
178. Emmanuel Paradis. Analysis of Phylogenetics and Evolution with R. 2006 Springer Science+Business Media. — 221 p.
179. Rowe, Philip.Essential statistics for the pharmaceutical sciences. 2007 John Wiley & Sons Ltd. — 287 p.
180. Antonella Bacchieri Giovanni Della Cioppa. Fundamentals of Clinical Research Bridging Medicine, Statistics and Operations. Springer-Verlag Italia, Milano 2004. — 343 p.
181. Tutorials in Biostatistics. Volume 1: Statistical Methods in Clinical Studies. Edited by R. B. D'Agostino, 2004 John Wiley & Sons Ltd. — 453 p.
182. Tutorials in Biostatistics. Volume 2: Statistical Modelling of Complex Medical Data. Edited by R. B. D'Agostino, 2004 John Wiley & Sons Ltd. — 486 p.
183. Biostatistics A Methodology for the Health Sciences. Gerald van Belle [et al.]— 2nd ed. 2004 by John Wiley & Sons. — 564 p.
184. Tutorials in Biostatistics. Volume 1: Statistical Methods in Clinical Studies. Edited by R. B. D'Agostino, 2004 John Wiley & Sons Ltd. — 453 p.
185. Geert Verbeke, Geert Molenberghs. Linear Mixed Models for Longitudinal Data. 2000 Springer-Verlag New York, Inc. —568 p.
186. Methods of Multivariate Analysis. Second Edition. ALVIN C. RENCHER. 2002 by John Wiley & Sons, Inc. — 708 p.
187. Paul Gustafson. Measurement Error and Misclassification in Statistics and Epidemiology. 2004 by Chapman & Hall/CRC. — 193 p.
188. Biostatistical Methods. Edited by Stephen W. Looney. 2002 Humana Press Inc. — 214 p.
189. Scott Long J. Regression Models for Categorical Dependent Variables Using STATA. 2001 by Stata Corporation. — 288 p.
190. Sabine Landau and Brian S. EverittA Handbook of Statistical Analyses using SPSS. 2004 Chapman & Hall/CRC Press LLC. — 354 p.
191. Epidemiology and Medical Statistics Handbook of Statistics. Volume 27. General Editor C.R. Rao 2007. — 852 p.
192. Optimum Experimental Designs, with SAS. A.C. Atkinson. A.N. Donev, R.D. Tobias. Oxford University Press Inc., New York. — 511 p.
193. Jiming Jiang. Linear and Generalized Linear Mixed Models and Their Applications. 2007 Springer Science Business Media, LLC. — 257 p.
194. Bevington, Philip R., Data Reduction and Error Analysis for The Physical Sciences. McGraw-Hill, Inc. — 320 p.
195. Longitudinal data analysis / Donald Hedeker, Robert D. Gibbons. 2006 by John Wiley & Sons, Inc. — 337 p.
196. Environmental and Health Risk Assessment and Management. Principles and Practices. 200 Springer. — 478 p.
197. Applied nonparametric statistical methods.—3rd ed. I.P. Sprent, N.C. Smeeton. Chapman & Hall/CRC. 2001. — 463 p.
198. Takeuchi Y., Iwasa K. Sato (Eds.) Mathematics for Life Science and Medicine. Springer-Verlag. 2007. — 227 p.
199. Stephen C. Newman. Biostatistical Methods in Epidemiology 2001. John Wiley & Sons, Inc. — 382 p.
200. Randomization in Clinical Trials Theory and Practice William F. Rosenberger., John M. Lachin. 2002. John Wiley & Sons, Inc. — 259 p.
201. Stephen Senn, Vic Barnett. Data Monitoring Committees in Clinical Trials. A Practical Perspective. 2002 John Wiley & Sons Ltd. — 191 p.
202. Michael A., Proschan K.K. Gordan Lan, Janet Turk Wittes. Statistical Monitoring of Clinical Trials. A Unified Approach. 2006 Springer Science+Business Media, LLC. — 258 p.
203. Ton J. Cleophas, Statistics Applied to Clinical Trials, Third Edition. — 2006 Springer. — 366 p.

Предметный указатель

Учебное издание

Петри М.А., **Сэбин** К.

НАГЛЯДНАЯ МЕДИЦИНСКАЯ СТАТИСТИКА

Перевод с английского под редакцией В.П. Леонова
2-е издание, переработанное и дополненное

Подписано в печать 25.08.2008. Формат 60×90 $^1/_8$. Бумага офсетная.
Печать офсетная. Объём 21 п.л. Тираж 1500 экз. Заказ 2051

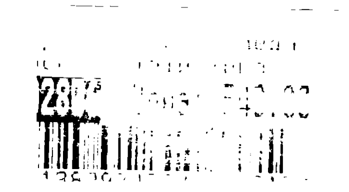
Издательская группа «ГЭОТАР-Медиа».
119435, Москва, ул. М. Пироговская, 1а;
тел.: (495) 921-39-07, факс: (499) 246-39-47,
e-mail: info@geotar.ru, http://www.geotar.ru

Отпечатано в ЗАО «Типография Полиграфист».
410031, г. Саратов, ул. Волжская, 28.

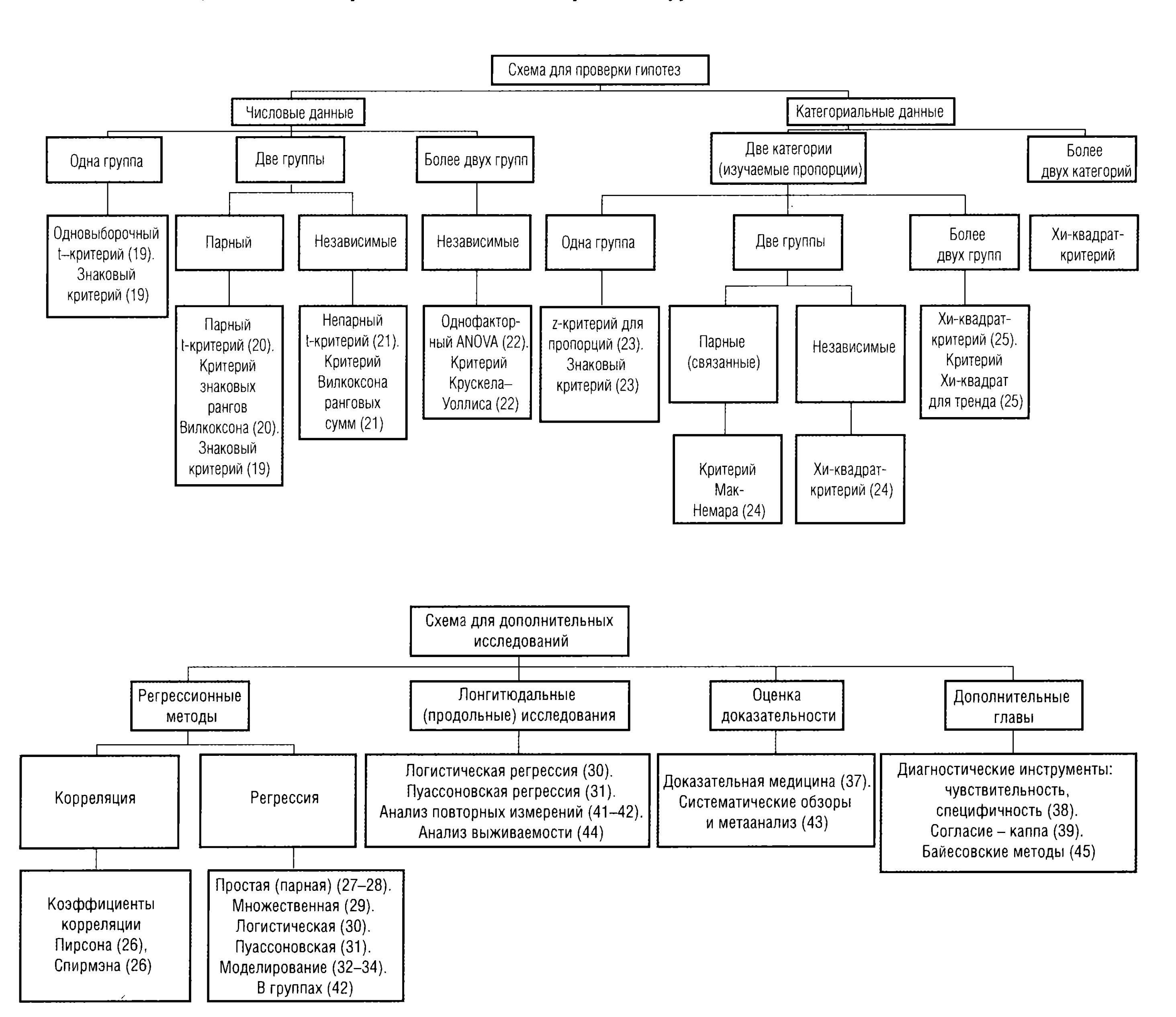





ИЗДАТЕЛЬСКАЯ ГРУППА
«ГЭОТАР-Медиа»



СХЕМЫ, УКАЗЫВАЮЩИЕ СООТВЕТСТВУЮЩИЕ МЕТОДЫ В РАЗЛИЧНЫХ ОБСТОЯТЕЛЬСТВАХ*



* Номера соответствующих глав приведены в скобках.

ОТЛИЧИТЕЛЬНЫЕ ОСОБЕННОСТИ СЕРИИ

- Краткость: основной материал по дисциплине в небольшом объёме.
- Наглядность: множество примеров и иллюстраций.
- Эффективность: студент готовится к экзамену на «отлично», а врач получает современное представление о предмете.

Цель второго издания книги, как и первого, — донести до читателя основные понятия и принципы медицинской статистики, которые довольно широко применяют зарубежные медики и биологи. Книга содержит не только необходимую теоретическую часть, но и в доступной форме даёт практическое описание того, как можно применять статистические методы в реальных клинических исследованиях. Для усвоения изложенного материала читателю вполне достаточно знаний по математике в объёме школьного курса.

Учебное пособие предназначено для студентов и аспирантов медицинских вузов, биологических факультетов университетов, врачей, исследователей-клиницистов и всех тех, кто интересуется применением статистики в медицине и биологии.

В серии вышли книги:

Наглядное акушерство и гинекология

Наглядная анатомия

Наглядная гастроэнтерология

Наглядная гематология

Наглядная иммунология

Наглядная кардиология

Наглядная неврология

Наглядная органическая химия

Наглядная офтальмология

Наглядная пульмонология

Наглядная фармакология

Наглядная хирургия

Наглядная эндокринология

Планируются к выпуску:

Наглядная медицинская генетика

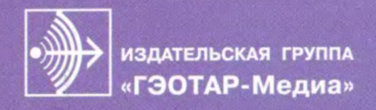
Наглядная микробиология и инфектология

Наглядная нефрология

Наглядная педиатрия

Наглядная физиология

Посетите наш сайт в Интернете: **www.geotar.ru**




ISBN 978-5-9704-0914-5
9 785970 409145